

OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OC NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12.
 RX MEDLINE-92114808; PubMed-1840644;
 RA "Sequencing, mutational analysis, and transcriptional regulation of
 RT the Escherichia coli htrb gene."
 RL Mol. Microbiol. 5:2285-2292(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-97061202; PubMed-8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kani K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horinouchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map."
 RL DNA Res. 3:137-155(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE-21074935; PubMed-11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lam A., Dimalanta E.T., Potamodis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blatter F.R.;
 RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7."
 RL Nature 409:529-533(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE-21156231; PubMed-11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
 RA Tida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12."
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RP CHARACTERIZATION.
 RX MEDLINE-91100364; PubMed-1846149;
 RA Karow M., Fayet O., Cegliska A., Zieglerhofer T., Georgopoulos C.;
 RT "Isolation and characterization of the Escherichia coli htrb gene,
 whose product is essential for bacterial viability above 33 degrees C
 in rich media."
 RL J. Bacteriol. 173:741-750(1991).
 RN [7]
 RP FUNCTION, AND CHARACTERIZATION.
 RX MEDLINE-90368739; PubMed-2203778;
 RA Brozek K.A., Raetz C.R.H.;
 RT "Biosynthesis of lipid A in Escherichia coli. Acyl carrier protein-
 dependent incorporation of laurate and myristate."
 RL J. Biol. Chem. 265:15410-15417(1990).

CC -1- FUNCTION: ACYLATES THE INTERMEDIATE (KDO)2-LIPID IVA TO FORM
 CC (KDO)2-(LAUROYL)-LIPID IVA. HAS 10 FOLD SELECTIVITY FOR LAUROYL-
 CC ACP OVER MYRISTOYL-ACP.
 CC -1- PATHWAY: LIPOPOLYSACCHARIDE CORE BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (potential).
 CC -1- INDUCTION: IS EXPRESSED AT ALL TEMPERATURES, BUT ACCUMULATION OF
 CC HTRB TRANSCRIPTS SLIGHTLY DECLINE WITH RAISING TEMPERATURE. THUS,
 CC ITS EXPRESSION IS NOT INDUCED BY HEAT SHOCK.
 CC -1- SIMILARITY: BELONGS TO THE HTRB/MSRB FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X61000; CAA43317.1; -
 CC DR EMBL; AE000206; AAC74138.1; -
 CC DR EMBL; D90742; BAA35852.1; -
 CC DR EMBL; D90743; BAA35863.1; -
 CC DR EMBL; AE005316; AAG55800.1; -
 CC DR EMBL; AP002555; BAB34855.1; -
 CC DR PIR; S16888; S16888.
 CC DR Ecogene; EG10464; htrb.
 CC KW Lipopolysaccharide biosynthesis; Transferase; Acyltransferase;
 CC Transmembrane; Inner membrane; Heat shock; Complete proteome.
 CC FT TRANSMEM 17 37 POTENTIAL.
 CC FT TRANSMEM 209 229 POTENTIAL.
 CC SQ SEQUENCE 306 AA; 35407 MW; AEA70A5EB10653B5 CRC64;
 CC -----
 CC alignment_scores: 215.00 Length: 274
 CC Quality: 1.303 Gaps: 9
 CC Percent Similarity: 60.219 Percent Identity: 29.197
 CC -----
 CC alignment_block:
 CC US-09-303-518D-569 x HTRB_ECOLI ..
 CC Align seg 1/1 to: HTRB_ECOLI from: 1 to: 306
 CC 88 CTGCCCGCTTCTGCTCGACACGCTGGGAACCGGCTGGACATCTGCG 137
 CC ||||| :||| ||||| ||||| |||||
 CC 32 LeupotyProVal11leTyArgleuglycyseglyLeuGlyLysLeuAl 48
 CC 138 GTTTTACCTTTTAAAGGAAGCGCGCGCATCTC...GCCAATATGC 184
 CC ||||| :||| ||||| ||||| |||||
 CC 48 aleuAlpHeMeLlyArg...ArgAlaLyslleValHisArgAsnLeug 64
 CC 185 GT.....CAGCGACGATGATATCCGACCCCAAAACGGTCAAA 222
 CC :||| :||| :||| :||| :||| :|||
 CC 64 luleucysPheProglumetserglIngluArgArgylsMetValVal 80
 CC 223 GCCGTTTTCGGAACGCAAAAGCGGCTTGGACATCTGCCCCCGCTT 272
 CC ||||| :||| ||||| ||||| |||||
 CC 81 LysAsnPhelGluSerValGlyMetGlyLeuMetGluThrGlyMetLarr 97
 CC 273 TTTTCAGAAACCGGACATAGAACATGTTCAAAAGCGGTACACGGCT 322
 CC :||| :||| :||| :||| :||| :|||
 CC 97 pPhE...TriProAspArgArglleAlaArgTrpThrGluValIleGlyw 113
 CC 323 GGGACATGTGACGACGCTTTGGACAAACAGAGGCGCTATTCATC 372
 CC ||||| :||| :||| :||| :||| :|||
 CC 113 etGluHisIleArgspValGlnAlaGlnLysArgGlylleLeuLeuVal 129
 CC 373 ACGCGGCAATCGCGGACGATTTGGCGGACGCTACATCATCAGCAGCA 422
 CC ||||| :||| :||| :||| :||| :|||
 CC 130 GlylleHisPheLeuThrLeuGluLeuGlyAlaArgGlnPheGlyMetG 146
 CC 423 GCTTCGCTCCCGCTGACCGCCATGTACAAACCGCGCAAAATCAAGCA 472

```

146 ngluProlyl.....ileglyValTyrArgProAsnAspAsnProLeu 161
473 TAGACAAATATCATGACGGGAGGAGGTTGGCGCAAGAAAAACCGCG 522
161 leAspTrpLeuGlnThrTrpValArgLeuAra.....SerAspLysSer 175
533 CCTACACATATCAAGGGTCAACAAATCATCAAGCCCTGCGTGGG 572
176 MetLeuAspArgLysAspLeuLysGlyMetIleLysAlaLeuLysGly 192
573 CGAACACCATCGCTGCGCCGACCGCTCCCTCCCTCAAGAGGCG 622
192 ygluValAlaTrpTyrAlaProAspHisAspTyrGlyProArg..... 206
623 GGAAGGCGTATGGTGGTATTTCTTCCG...AAACCTGCCTATACATG 669
207 ..SerSerValPheValProLeuPheAlaValGluGlnAlaThrThr 222
670 ACGCTGGCGCAAAATTTGGACACGCTCAAGGCGTGAACCCCTGTTT 719
223 ThrGlyThrTrpMetLeuAlaArgMetSerGlyAlaCysLeuValPro 239
720 CTGCTGGCAAGCGCTGCTGCGGACAAAGTTTGATTTGACATCCGCC 769
239 evalProArgTyrGlyProAspGlyLysGlyTyrGlnLeuLeuMetLeu 256
770 CCGTCCAGGGGGAATTTGACGCGCAAGGCCCAT...GATCGCGCGTG 816
256 roProGluCysSerProProLeuAspAspAlaGluThrThrAlaAlaTrp 272
817 TTCACGCGCAATGCCGAATATTTGATACGCCGTTTCCGACGCAATCT 866
273 MetAsnLysValValGluLysCysIleMetMetAlaProGluGlnTyr 289
867 GTTATGTACACCGCTACAA 888
289 tTrpLeuHisArgArgPheLys 296

seq_name: SwissProt_40:DDG_ECOLI

seq_documentation_block:
ID DDG_ECOLI STANDARD: PRT: 306 AA.
AC P76522: P76949:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 16-OCT protein.
DE DDG OR B2378.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,

```

```

RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Oshiba T., Oyama S., Saito N., Sempel G., Saton Y., Sivasubaram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE HTRB/MSB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U49787; AAB6658.1; -
DR EMBL: AE000326; AAC75437.1; ALT_INIT.
DR EMBL: D90868; BAA16248.1; -.
DR EcoGene: EG1901; ddg.
KW Transmembrane; Inner membrane; Transferase; Acyltransferase;
KW Complete proteome.
FT TRANSMEM 20 40 POTENTIAL.
FT TRANSNEM 125 145 POTENTIAL.
SQ SEQUENCE 306 AA: 35493 MW: 58327F3D969E7B6 CRC64;

```

```

alignment_scores:
Quality: 189.50 Length: 294
Ratio: 1.199 Gaps: 12
Percent Similarity: 53.741 Percent Identity: 26.190

```

alignment_block:

US-09-303-518D-569 x DDG_ECOLI ..

Align seg 1/1 to: DDG_ECOLI from: 1 to: 306

```

88 CTGCGCGCTTCTGTCTGTCACACGCTGGGAAACGGCTGGACATCTGCG 137
||||| ||| ||||||||| |||||
32 LeuProTyrProValLeuLysPheLysGlyThrArgIleGlyAlaMetAl 48
138 GTTATACCTTTAAAG..... 153
48 argProPheLeuLysArgArgGluSerIleAlaArgLysAsnLeuGluL 65
154 .....GAAGACCGCGCGCGCATCGTCCG... 177
65 euCysPheProGlnHisSerAlaGluGluArgGluLysMetIleAlaGlu 81
178 AATATCGCTGACGACGATGAATCCGACCCCAAGCGTAAAGCCGT 227
||||| |||||
82 AsnPheArgSerLeuGlyMet..... 88
228 TTTTGGGAAAGCGCAAAAGCGGTTTGAATTCGCCCGCGGTTTTC 277
89 .....AlaLeuValGluThrGlyMetAlaTrpPheT 99
278 GAAACCGGAGACATAGAACATGTTTCAAGCGGTACAGCGCGGGA 327
::: ||| |||||
99 rProAspSerArgValArgLysTrpPheAsp...ValGluLysLeuAsp 114
328 CATGTGACAGCGCTTTGGACAAACAGAGGCTGATTCATCAGCGCC 377
||||| |||||
115 AsnLeuLysArgAlaGlnMetGlnAsnArgGlyValMetValGlyLys 131
378 GCACATCGGCACACTACGATTTGGCGGACGCTACATCAGCAGCGCTTC 427
||||| |||||
131 HisPheMetSerLeuGluLeuGlyGlyArgValMetGly.....Leuc 146

```



```
558 AGCCCTGCGTGGGGAAGCAACCATGCTCCGCCGACAGTCCCT 607
      : : : : : : : : : : : : : : : : : : : : : :
202 RHSAVALARGLYGLUMETGLYTYRTHLEUPROASP..... 215
      : : : : : : : : : : : : : : : : : : : :
608 CCCCTCAAGAGCGGG.....GAAGCGTGTGGGTGATTTCTTGGC 651
      : : : : : : : : : : : : : : : : : : : :
216 ....GLUSPHEGLYALAGLUGINSERVALPHEVALASPHHEGLY 230
      : : : : : : : : : : : : : : : : : : : :
652 AAACCTGCTTATACCATGACGCTGGCGGCAAAATGGCACAGTC...AA 698
      : : : : : : : : : : : : : : : : : : : :
231 THRTYRLYALATHRLEUPROGLYLEUASNLYSMETALALYSLEUSERLY 247
      : : : : : : : : : : : : : : : : : : : :
699 AGCGGTGAAGAACCCGTGTTTCTGCTGCGAACCCGCTGCGGAGCAAG 748
      : : : : : : : : : : : : : : : : : : : :
247 SAIAVALVALIIEPROMETLPHROARGTYRASNLAGLTHRGLYLS. 263
      : : : : : : : : : : : : : : : : : : : :
749 GTTGGATTGGACATCGCCGCCGCAAGGGAATTGACGGCACA 798
      : : : : : : : : : : : : : : : : : : : :
264 ..TYGLUMETGLIULEHISPROALAMET...ASNLEUSERSPASPRO 278
      : : : : : : : : : : : : : : : : : : : :
799 GCCCATGATGCCCGCGTGTTCACCGCAATGCGCAATTTGGATACGGCG 848
      : : : : : : : : : : : : : : : : : : : :
279 GLUGINSERVALARGALAMETASNGLUGIULEGILUSERPHEVALTHPR 295
      : : : : : : : : : : : : : : : : : : : :
849 TTTTCCGACGAGATCTGTTTATG 873
      : : : : : : : : : : : : : : : : : : : :
295 CALAPROGLUGINTYRVALTPRIE 303
      : : : : : : : : : : : : : : : : : : : :
seq_name: SwissProt_40:HTRB_HAEIN
seq_documentation_block:
ID HTRB_HAEIN STANDARD; PRT; 311 AA.
AC P45239; Q48045;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lipid A biosynthesis lauroyl acyltransferase (EC 2.3.1.-) (Heat shock
protein B).
GN HTRB OR MAAM OR H11527.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxId=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA McKenney K., Sulton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhman J.L., Geoghagen N.S.M.,
RA Greim C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT 'Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.';
RL Science 269:496-512(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NH1 2019;
RX MEDLINE=96070820; PubMed=7592970;
RA Lee N.-G., Sunshine M.G., Engstrom J.J., Gibson B.W., Apicella M.A.,
RT 'Mutation of the htrb locus of Haemophilus influenzae nontypable
RT strain 2019 is associated with modifications of lipid A and
RT phosphorylation of the lipopolysaccharide.';
RL J. Biol. Chem. 270:27151-27159(1995).
CC -1- FUNCTION: ACYLATES THE INTERMEDIATE (KDO)2-LIPID IVA TO FORM
CC (KDO)2-(LAPROYL)-LIPID IVA (BY SIMILARITY).
CC -1- PATHWAY: LIPOLYSACCHARIDE CORE BIOSYNTHESIS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
```

```
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE HTRB/MSBB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL; U32828; AAC3173.1; ALT_INIT.
DR EMBL; U17642; AAC43515.1; -.
DR TIGR; H11527; -.
KW Lipopolysaccharide biosynthesis; Transferase; Acyltransferase;
KW Transmembrane; Inner membrane; Heat shock; Complete proteome.
FT TRANSMEM 18 39
FT POTENTIAL.
FT TRANSMEM 210 231
FT E -> K (IN REF. 2).
FT CONFLICT 57 57
FT V -> T (IN REF. 2).
FT CONFLICT 81 81
FT T -> V (IN REF. 2).
FT CONFLICT 270 270
FT A -> V (IN REF. 2).
FT CONFLICT 274 274
FT DITI -> GISO (IN REF. 2).
FT CONFLICT 287 290
FT K -> N (IN REF. 2).
FT CONFLICT 305 305
SQ SEQUENCE 311 AA; 36152 MW; F4250F7B1B5B6845 CRC64;
```

```
alignment_scores:
Quality: 148.50 Length: 287
Ratio: 0.946 Gaps: 11
Percent Similarity: 54.704 Percent Identity: 25.087
```

alignment_block:

US-09-303-518D-569 x HTRB_HAEIN ..

Align seq 1/1 to: HTRB_HAEIN from: 1 to: 311

```
64 GCGCTGCTCAATAGCCCTGCTGCGCGCTTCCGTGCGACACGCG 113
      : : : : : : : : : : : : : : : : : : : :
26 ALALETPRARGSERILEUCYSLEPROTYRPROLLEUARGHIS11 42
      : : : : : : : : : : : : : : : : : : : :
114 GGGAAACGGGCTCGGACATCTGGCGTTTACCTTTTAAAGAACGGCG 163
      : : : : : : : : : : : : : : : : : : : :
42 EGLYHISGLYPHEGLYTRLEUPHESEHISLEULYVALGLUGLARG 59
      : : : : : : : : : : : : : : : : : : : :
164 CGCGCATCGTGGCCATATGCGTCAGGACGCGATGCCGAC...CCC 210
      : : : : : : : : : : : : : : : : : : : :
59 RGLAALALALALALARGARGASNLEUCIDLEUCYSRHEPROASPME 75
      : : : : : : : : : : : : : : : : : : : :
211 AAACGGTCAAGACCGTGTGTTGGGAA.....ACGGCANA 245
      : : : : : : : : : : : : : : : : : : : :
76 GLASNGLARGLVALIILEUGINGLUSNLEUARGSERVALIYLE 92
      : : : : : : : : : : : : : : : : : : : :
246 AGCGGTGTTGGAACCTGCCCGCGTGTTCAGAAAACCGAAGACATAG 295
      : : : : : : : : : : : : : : : : : : : :
92 TALAILELLEGLTHRGILMETALTRPHERPESRASPERSERARGI 109
      : : : : : : : : : : : : : : : : : : : :
296 AAACATGTTCANAGCGGTACACGCTGGGAACATGTCAGCAGGCTTG 345
      : : : : : : : : : : : : : : : : : : : :
109 YLSYSTRPSERLYS...VALGLUGLYLEUHSITRYLEUYSGLU... 122
      : : : : : : : : : : : : : : : : : : : :
346 GACAAACAGGAGCGTCTATTCATCACGCGGACATCGGACGTACGA 395
      : : : : : : : : : : : : : : : : : : : :
123 ASNGILYASPLYLILEVALLEUVALIGLYVALHISPHLEUTHRLE 139
      : : : : : : : : : : : : : : : : : : : :
396 TTGGGCGGACGCTACATAGCCAGCAGCTTCCGTCGTCAGCGCA 445
      : : : : : : : : : : : : : : : : : : : :
139 ULEUGLYALARGILELLEGLY.....LEUHSITRPGYLLILEGLY 154
      : : : : : : : : : : : : : : : : : : : :
446 TGTACAAACCGCGGAATCAAGCGATGACAAATATCGAGGCGGC 495
      : : : : : : : : : : : : : : : : : : : :
154 ALTYRTPROASNPSASNPROLEULEUASPTPLUGINTHRGLNGLY 170
```

```

496 AGCGTGGC...GGCAAGAGAAAAAGCGGCTACGACATACAGGGGT 542
|||||.....|
171 ArgLeuATGSerAsnLysASPmetLeuAspArgLysAspLeuArgGly.. 186
543 CAACAATATCATCAAGCCCTGCGTGGGCGAAGCAACATCTCTGTC 592
|||||.....|
187 .....MetLeuLysAlaLeuATGHisGluGluThrIleThrPyrLap 201
593 CCGACGACGTCCTCCCTCCCTCAAGAGCGGG.....GAAGCGGTATGC 636
|||||.....|
201 rOASpHis.....AspTYrGlyArgLysAsnAlaValPhe 212
637 GTGATTTCTTGGCAACCTGCGCATACATGATGAGCGGGCGGA.... 681
|||||.....|
213 ValProPhePheAlaValProAspThrCysThrThrThrGlySerTYT 229
682 .AAATGGCACACGTCGCAAGGCGTGAAGAACCTGTTTGTGCGGCAAC 730
|||||.....|
229 rLeuLeuLysSerSerGlnAsnSerLysValIleProPheAlaProLeuA 246
731 GCCTGCGCGGCGGCAAGCTTTCGATTGCAATC...CGCCCGTCCAA 777
|||||.....|
246 rGAsnLysAspLysArgLysGlyThrValSerLieserAlaProValAsp 262
778 GGGGATTTGAACGGCGCAAGCCCATGATGCGCGGTGTTCAACCGCAA 827
|||||.....|
263 PheThrAspLeuGlnAspLysGluThrAlaIleAlaValArgMetAsnGlnI 279
828 TGCCCAATATTGGATACCGCGCTTTCACGACGATCTGTTATGTACA 877
|||||.....|
279 eValAluLysGluIleMetLysAspLierThrIleYrMetTrpLeuHisA 296
878 ACCGCTACAA 888
296 rGArgPheLys 299
seq_name: SwissProt.40:MSBB_ECOLI
seq_documentation_block:
ID MSBB_ECOLI STANDARD; PRT; 323 AA.
AC P24205;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lipid A biosynthesis (KDO)2-(Lauroyl)-lipid IVA acyltransferase
DE (EC 2.3.1.-)
OS MSBB OR B1855.
CN Escherichia coli.
CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92121107; PubMed=1732206;
RA Karow M., Georgopoulos C.;
RT "Isolation and characterization of the Escherichia coli msbB gene, a
RT multicopy suppressor of null mutations in the high-temperature
RT requirement gene htrb.";
RL J. Bacteriol. 174:702-710(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93015688; PubMed=1356966;
RA Engel H., Smitk A.J., van Wijngaarden L., Keck W.;
RT "Moren-metabolizing enzymes from Escherichia coli: existence of a
RT second lytic transglycosylase";
RL J. Bacteriol. 174:6394-6403(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9276503;
RA Baltner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

```

```

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sampei G., Seki Y., Sivasubram S., Tagami H.,
RA Takeda J., Takemoto K., Mada C., Yamamoto Y., Horiiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
RN [5]
RP FUNCTION, AND CHARACTERIZATION.
RX MEDLINE=97256743; PubMed=909672;
RA Clementz T., Zhou Z., Raetz C.R.H.;
RT "Function of the Escherichia coli msbB gene, a multicopy suppressor
RT of herb knockouts, in the acylation of lipid A. acylation by msbB
RT follows laurate incorporation by Htrb.";
RL J. Biol. Chem. 272:10353-10360(1997).
CC -1- FUNCTION: TRANSFERS MYRISTATE OR LAURATE, ACTIVATED ON ACP, TO THE
CC LIPID IVA MOIETY OF (KDO)2-(LAUROYL)-LIPID IVA. DECANOYL,
CC PALMITOYL, PALMITOLEOYL, AND (R)-3-HYDROXYMYRISTOYL-ACP ARE POOR
CC ACYL DONORS. FUNCTIONS OPTIMALLY AFTER LAURATE INCORPORATION BY
CC HTRB HAS TAKEN PLACE. ACYATES (KDO)2-(LAUROYL)-LIPID IVA ABOUT
CC 100 TIMES FASTER THAN (KDO)2-LIPID IVA. DISPLAYS A PREFERENCE FOR
CC MYRISTOYL-ACP OVER LAUROYL-ACP.
CC -1- PATHWAY: LIPOPOLYSACCHARIDE CORE BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -1- SIMILARITY: BELONGS TO THE HTRB/MSBB FAMILY.
CC -1- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE THE MEMBRANE-BOUND
CC LYTIC MOREN TRANSGLYCOSYLASE (MLT).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M77039; AAA24181.1; -.
DR EMBL: M87660; AAA96706.1; -.
DR EMBL: AE000279; AAC74925.1; -.
DR EMBL: D90828; BAA15663.1; -.
DR PIR: A42608; A42608.
DR EcGene; BG10614; msbB.
KW Lipopolysaccharide biosynthesis; Transferase; Acyltransferase;
KW Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 23 43 POTENTIAL.
FT TRANSMEM 85 105 POTENTIAL.
FT TRANSMEM 133 153 POTENTIAL.
FT SEQUENCE 323 AA; 37410 MW; 94DAF38A57D20CD CRC64;

```

```

alignment_scores:
  Quality: 127.50      Length: 297
  Ratio: 0.867        Gaps: 11
Percent Similarity: 49.495      Percent Identity: 23.232

```

alignment_block:

us-09-303-518d-569 x MSBB_ECOLI ..

Align seq 1/1 to: MSBB_ECOLI from: 1 to: 323

```

64 GCGCTGCTCAATGCGCTCTCCGTCGCTGCTGTCGACACGCT 113
||| ::: :|:::|::| ||| :::

```

```

31 AlaAlaMetAlaGlyLeuAlaLeuThrProProLysPheArgAspProI 47
114 GGGAAACCGCGTCGACATCTGGCGTTTACCTTTAAAGAAAGACCGCG 163
    |||||
47 eLeuAlaIaGlyLeuGlyArgPheAlaIaGlyArgLeuGlyLysSerArgA 64
164 CGCGCATCGTCCCAATATG..... 183
64 rgrAlaLeuIleAlaLeuSerLeuCysPheProGluArgSerGluA 80
184 ...CGTCAGGAGCGATGATCCCGACCCCAAAACGGTCAAAAGCGTTT 230
    |||||
81 GluArgGluAla.....IleValAspGluMetPh 90
231 TCGCGAAACGGCAAAAGCGGTTTGGAACTTGCCTCCCGCGTTTTCAGAA 280
    |||||
90 eAlaThrAlaProGlnAlaMetAlaMetAlaGluLeuAlaIleArgG 107
281 AACCGAAGACATAGAAACATGTTCAAGCGGTACAGCGGTGGAGACAT 330
    |||||
107 LProGluLysIleGlnProArgValAsp...TrpGlnGlyLeuGluIle 122
331 GTCGAGCGCGCTTTGGACAAACACAGAGCGCTGATTATCATCAGCGCGA 380
    |||||
123 ILeGluGluMetArgArgAsnAsnGluLysValIlePheLeuValProH 139
381 CATCGGACGCTACGATTGGCGGACGCTACATCAGCGACGAGCTTCGCT 430
    |||||
139 selyTrpAlaValAspIleProAlaMetLeuMetAlaSerGln...GlyG 155
431 TCCCGCTGACCGCCATGTACAAACCGCGAAATCAAGCATAGACAAA 480
    |||||
155 InLysMetAlaIaMetPheIleAsnGlnGlyAsnProValPheAspArg 171
481 ATCATGCAAGCGCGGCGGCTTCGCGCAAGAAACCGCGCTACAG 530
    |||||
172 ValTrpAsnThrValArgArgArgPheGlyArgLeuHisAlaArgAs 188
531 CATACAAAGGGGTCAAAACATATCAAAAGCGGTGGCGGCAAGCA 580
    |||||
188 n...AspGlyIleLysProPheIleGlnSerValArgGlnGlyTrpG 204
581 CCATCGCTCGTCC...GACCAAGTCCCTCCCTCAAGAAAGCGGG 624
    |||||
204 LYTTrpTrpLeuProAspGlnAspHisGlyProGlnHisSerGlu..... 218
625 GAAGCGGTATGGTATTTCTCGGCAAAACCGCTATACCATGACGCT 674
    |||||
219 .....PheValAspPhePheAlaThrTrpLysAlaThrLeuProAl 232
675 GCGGCAAAATTTGGCACACGTCAAAAGCGGTGAAAAACCTGTTTCTGCT 724
    |||||
232 aIleGlyArgLeuMetLysVal.....CysA 241
725 GCGAAGCGCTG.....CCTGGCGGACAAAGT..... 750
    |||||
241 rGAlaIaGlyValAlaProLeuPheProIleTrpAspGlyLysThrHisArg 257
751 TTGCATTTGCACATCGCCCGCTCCAAAGGGGA...TTGAAGCGGACAA 797
    |||||
258 LeuThrIleGlnValArgProPheMetAspAspLeuLeuGluAlaAspAs 274
798 AGCCCATGATGCCCGCTGTTCAACCCCAATGCCGAATATTGATACGCC 847
    |||||
274 PhISThrIleAlaArgArgMetAsnGluGluValGluIlePheValGly 291
848 GTTTTCGACGACATATCTGTTTATGTACAAACCGCTACAAA 888
    |||||
291 roArgProGluGlnIleTrpThrIleLeuLysLeuLeuLys 304

```

seq_name: SwissProt_40:YGY3_HAISO
seq_documentation_block:

```

ID YGY3_HAISO STANDARD; PRT; 437 AA.
AC P21561;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 50.6 kDa protein in the 5' region of GYRA and GYRB (ORF
DE 3).
OS Haloferax sp. (strain Aa 2.2).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloferax.
OX NCBI_TaxID=2254;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91100352; PubMed=1846146;
RA Holmes M.L., Dyall-Smith M.L.;
RT "Mutations in DNA gyrase result in novobiocin resistance in
RT halophilic archaeobacteria."
RL J. Bacteriol. 173:642-648(1991).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC
CC EMBL: M3873; -; NOT_ANNOTATED_CDS.
DR PIR: C39135; C39135.
KW Hypothetical protein.
SQ SEQUENCE 437 AA; 50626 MW; B5B99A2AF3892BEF CRC64;

```

alignment_scores:
Quality: 115.00 Length: 296
Ratio: 0.935 Gaps: 16
Percent Similarity: 41.554 Percent Identity: 26.014

alignment_block:

US-09-303-518D-569 x YGY3_HAISO ..

Align seq 1/1 to: YGY3_HAISO from: 1 to: 437

```

108 CAGCGTGGAAACCGCGCTCGACATCGGCGGTTTACCTTTAAAGGANG 157
    |||||
119 HisAlaGly...AspArgArgAlaProGlyVal.....AspSerArg 131
158 ACCGCGCGCGCATGTCGCCAATATGCGTCAGCGACGATGAATCC...C 204
    |||||
131 gLeuArg...GlnGlnHisGlnHisProArgGlyArgHisAlaSerAspA 147
205 GACCCCAAAACGGT.....CAAGC 224
147 rGValGlnAspGlyAlaHisProAlaArgGlnArgLeuArgGluGlnPro 163
225 CGTTTTTGGGAAACCGCAAAAGCGGTTTGAACCTTGCCTCCCGCTTT 274
    |||||
164 ArgHisAlaGlyArgProArgArgArg.....GlnProProArgArgGln 178
275 TCGAAACCGGACATAGAAACATGTTCAAAAGCGGTACAGCGGTG 324
    |||||
178 yArgSerArgGlyThrHisArgArgHisLeuArgGlnAlaProAlaProAl 195
325 GAACATGTGACAGCGCTTTGGCAAAACGAAAGGCGTCTTTCATCAC 374
    |||||
195 lValArgGlyProAspGlnAspGlnAlaArgGlu..... 206
375 GCGGCAATCGGACGCTACGATTTGGCGGCGAGCGTACACGACGAC 424
    |||||
207 .....PheArgGlyProArgHisArgArgGluArgHisProProThrAl 221
425 TTCGCTTCCCGCGTACGCGCATGACAAACCGCGCAAAATCAAGCGATA 474
    |||||
221 a.....ArgAspValLeuArgGlyLysProGlnHisGlyA 233

```



```

724 mlaserPhearsgluleuthrCysHisCysGlyAlaserValIleTyr 741
579 AACCATGCTCTGCCCGACAGTCCCT..... 607
741 roPvalProCysGlyThrArgProGluCysThrGlnThrCysAla 757
608 .....CCCCCAAGAGCGGGGAGGCGTATGG 636
758 ArgValHisGluCysAspHisProValTyrHisSerGlyHis..... 771
637 GTGATTTCTTGGCAACCTGCT.....ATACATAGCGT 674
772 .....SerGluGluLysCysProProCysThrPheLeuThrGlnLysT 786
675 GGGGCAAAATGTGCACAGTCACAGCGTCAAGAGCGTGT..... 718
786 rPcysMetGlyLysHisGluPheArgSerAsnIleProCysHisLeuVal 802
719 .....TGTGTGCGAACGCTGCTGCGCGA 744
803 AspIleSerCysGlyLeuProCysSerAlaThrLeuProCysGlyMetI 819
745 CAAGGTTTCATTTGCACATCCGCCCGCAAGGGAATGACGGCGA 794
819 sLys.....CysGlnArgLeuCysHisLysGlyLysLeuVala 833
795 CAAGCCCAAGATGCCCGCTGTTCACCGCAATGCCAATATGATGATC 844
833 spGluProCysLysGlnProCysThrThrProArgAlaAspCysGlyHis 849
845 GCCGTTTCCGACGAGTATCTTTA.....TGTACCAACGCGTCAAA 888
850 Pro.....CysMetAlaProCysHisThrSerTr 860
889 ATGC 892
860 cCys 861

seq_name: SwissProt_40:CRE1_NEUCR
seq_documentation_block:
ID CRE1_NEUCR STANDARD: PRT; 430 AA.
AC 059958:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE DNA-binding protein cre-1 (carbon catabolite repressor).
GN CRE-1.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RA de la Serna I., Tyler B.M.; to the EMBL/GenBank/DBJ databases.
RL Submitted (MAR-1998)
CC -1- FUNCTION: INVOLVED IN CARBON CATABOLITE REPRESSION. REPRESENTS THE
CC TRANSCRIPTION OF A NUMBER OF GENES BY BINDING TO A GC-RICH REGION
CC IN THEIR PROMOTER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: BELONGS TO THE CRE1/MIG GROUP OF C2H2-TYPE ZINC-
CC FINGERS PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: AF05464; AAC13555.1; -.
DR HSP: P07248; 1ARD.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF00096; Zf-C2H2_2.
DR PRINTS: PR00048; ZNCFINGER.
DR SMART: SM00355; ZNF_C2H2_2.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 2.
DR DNA-binding: Transcription regulation; Repressor; Zinc-finger;
KW Nuclear protein; Metal-binding.
KM ZN_FING 78 100 C2H2-TYPE.
FT ZN_FING 106 130 C2H2-TYPE.
FT ZN_FING 36 45 POLY-GUN.
FT DOMAIN 142 146 POLY-GUN.
SQ SEQUENCE 430 AA; 46954 MW; FDD30489A7D8085 CRC64;

```

```

alignment_scores:
  Quality: 110.50      Length: 338
  Ratio: 0.773         Gaps: 16
  Percent Similarity: 42.308      Percent Identity: 22.189

```

alignment_block:

US-09-303-518D-569 x CRE1_NEUCR ..

Align seg 1/1 to: CRE1_NEUCR from: 1 to: 430

```

18 CAGCGTGTTCCTCCCTTTCGAAACGCGCATGCAATCTGTTGACGCGC 67
   |||||:|||||
66 GlnAlaThrGluPro.....AlaAsnGluLeuProArgPr 77
68 TGTCTCAATGCTCTCTCCCTGCTGCGCTTC..... 98
   |:::|||||
77 oTylLysCysProLeuCysAspLysAlaPheHisArgLeuGlnHisGlnT 94
99 .....CTGTGCAACAGCTGGG...AAACGCGCTCGACATCTGGC 137
94 hArgHisIleArgThrHisThrGlyLysPro.....His 106
138 GTTTTACCTTTTAAAGAGACCGCGCGCATGTCGCAATATGCGTC 187
   ::|||
107 AlaCysGlnPheProGlyCysSerLysPheSerArgSerAlaPheIle 123
188 AGCGAGCATGATCCGACCGCCCAAAACGCTCAACCGCTTTTGGCGAA 237
   |||||:|||||
123 uThrArgHisSerArgLleHisSerAsnProAsnSerArgGlyAsnL 140
238 ACGGCAAAAGCGCG.....TTTGAACCTGCG 263
   ::|||:|||||
140 yGlyGlnGlnGlnGlnHisProLeuValHisAsnHisGlyLeuGln 156
264 CCGCGCGTTCAGAAACCGGA..... 287
   |||
157 ProAspMetMetProProGlyProLysAlaIleArgSerAlaProPr 173
287 ..... 287
173 oThrAlaMetSerSerProAsnValSerProPronHisSerTyrSerPro 190
288 .....AGCATAGA 296
190 yAsnPheAlaProSerGlyLeuAsnProTyrSerHisSerAlaGlySer 206
297 AACATGTTCAAAGCGGTACAGCGCTGGGAACATGTCACAGAGCTTGG 346
   ::|||:|||||
207 AlaGlySerGlnSerGlyProAspIleSerLeuLeuAlaArgAlaLagI 223
347 ACAAAACAGAACG.....GCTGCTATTCATACAGCGCGACATGCGCAGC 390
   |||||:|||||
223 yGlnValGlnArgAspGlyAlaAlaHisHisHisPheGlnProArgPheG 240
391 TACGATTTGGGCGAGCGTACATGACGACGAGCTTC.....GTT 431
   ::|||:|||||

```


798 1stYrGlyInThrGlyGlyGlyLeuSerHisGlnGlyInThr 814
 501 TCAGGCGAAGAAAGCCGCTACAGCATACA.....AGGGGTCA 544
 815 AspSerGlnGlyGlnAsnSerGlnInThrPHisArgThrAspSerGlnGly 831
 545 AACAAATCAT.....CAAGCCCTCGCTCGGCGCAACCAATCGTC 588
 831 nserPheHisPheAspGlnAlaGlyArgGlyGlnSerSerHisHisG 848
 589 CTGGCCGACACACGTCCTCCCTCCCAAGAGGGGAGG...CGTATG 635
 848 1yGlnThrAspArgGlnSerGlnSerHisGlyGlnSerGlnGln 864
 636 GGTGATTTCTTCGCAACCTCGCTATACATGACGCTGC..... 677
 865 GlyThrGlnAsnGlnGlnGlnGlnAsnArgHisSerLeuGlyThrAsp 881
 678GGCAAAATG 687
 881 gThrArgArgAspSerTyValGlnHisSerGlyArgSerGlyLeu 898
 688 GCACACGCTCAAGGCTGAACCCCTGTTTCTGTCGACAGCCCTGCC 737
 898 ergGlnGlnAsnSerArgGlnGlu.....Val 906
 738 TGGCGGACAGGTTTCGATTTGCACATCCGCGCTCCAAAGGGAATGA 787
 907 ArgGlnThrGlnSerGlnArgSerHisAspArgArgGlnGlnGln 923
 788 ACGGCGA 794
 923 nGlnGln 925

seq_name: SwissProt_40:AT7A_HUMAN

seq_documentation_block:
 ID AT7A_HUMAN STANDARD; PRT; 1500 AA.
 AC 004656; 000745; 000227;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Copper-transporting ATPase 1 (EC 3.6.3.4) (Copper pump 1) (Menkes
 disease-associated protein).
 GN ATP7A OR MNK OR MCL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RC TISSUE=Fibroblast;
 RX MEDLINE=93258410; PubMed=8490659;
 RA Vulpe C.D., Levinson B., Whitney S., Packman S., Gitlschier J.;
 RT "Isolation of a candidate gene for Menkes disease and evidence that
 it encodes a copper-transporting ATPase.";
 RL Nat. Genet. 3:7-13(1993).
 RN [2]
 RP ERRATUM.
 RA Vulpe C.D., Levinson B., Whitney S., Packman S., Gitlschier J.;
 RL Nat. Genet. 3:273-273(1993).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RX MEDLINE=9533177; PubMed=7607665;
 RA Tuemer Z., Vural B., Toennesen T., Chelly J., Monaco A.P., Horn N.;
 RT "Characterization of the exon structure of the Menkes disease gene
 using vectorette PCR.";
 RL Genomics 26:437-442(1995).
 RN [4]
 RP SEQUENCE OF 1-1447 FROM N.A. (ISOFORM 4).
 RX MEDLINE=96039257; PubMed=7490081;
 RA Dietrick H.A., Ambrosini L., Spencer J., Glover T.W., Mercer J.F.B.;
 RT "Molecular structure of the Menkes disease gene (ATP7A).";

RL Genomics 28:462-469(1995).
 RN [5]
 RP SEQUENCE OF 1-626 FROM N.A. (ISOFORM 4).
 RC TISSUE=Kidney;
 RX MEDLINE=93258397; PubMed=8490646;
 RA Chelly J., Tuemer Z., Toennesen T., Petterson A., Ishikawa-Brush Y.,
 RT Tommerup N., Horn N., Monaco A.P.;
 RT "Isolation of a candidate gene for Menkes disease that encodes a
 potential heavy metal binding protein.";
 RL Nat. Genet. 3:14-19(1993).
 RN [6]
 RP SEQUENCE OF 12-529 FROM N.A. (ISOFORM 4).
 RC TISSUE=Endothelial cells;
 RX MEDLINE=93258398; PubMed=8490647;
 RA Mercer J.F.B., Livingston J., Hall B., Paynter J.A., Begy C.,
 RA Chandrasekharappa S., Lockhart P., Grimes A., Blave M.,
 RA Stameniak D., Glover T.W.;
 RT "Isolation of a partial candidate gene for Menkes disease by
 positional cloning.";
 RT Nat. Genet. 3:20-25(1993).
 RN [7]
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RA Bird C.;
 RL Submitted (APR-1997) to the EMBL/Genbank/DBJ databases.
 RN [8]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RC TISSUE=Fibroblast, and Colon carcinoma;
 RX MEDLINE=99179550; PubMed=10079814;
 RA Harris E.D., Reddy M.C., Qian Y., Tiffany-Castiglioni E., Majumdar S.,
 RA Nelson J.;
 RT "Multiple forms of the Menkes Cu-ATPase.";
 RT Adv. Exp. Med. Biol. 448:39-51(1999).
 RN [9]
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RC TISSUE=Fibroblast;
 RX MEDLINE=98359731; PubMed=9693104;
 RA Reddy M.C., Harris E.D.;
 RT "Multiple transcripts coding for the menkes gene: evidence for
 alternative splicing of Menkes mRNA.";
 RT Biochem. J. 334:71-77(1998).
 RN [10]
 RP ALTERNATIVE SPLICING (ISOFORM 5), AND SUBCELLULAR LOCATION.
 RX MEDLINE=98133927; PubMed=9467005;
 RA Qi M., Byers P.H.;
 RT "Constitutive skipping of alternatively spliced exon 10 in the ATP7A
 gene abolishes Golgi localization of the menkes protein and produces
 the occipital horn syndrome.";
 RL Hum. Mol. Genet. 7:465-469(1998).
 RN [11]
 RP ALTERNATIVE SPLICING (ISOFORM 6).
 RC TISSUE=Neuroblastoma;
 RX MEDLINE=20427712; PubMed=10970802;
 RA Reddy M.C., Majumdar S., Harris E.D.;
 RT "Evidence for a menkes-like protein with a nuclear targeting
 sequence.";
 RL Biochem. J. 350:855-863(2000).
 RN [12]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=97227286; PubMed=9147644;
 RA Dietrick H.A., Adam A.N., Escara-Wilke J.F., Glover T.W.;
 RT "Immunocytochemical localization of the Menkes copper transport
 protein (ATP7A) to the trans-Golgi network.";
 RL Hum. Mol. Genet. 6:409-416(1997).
 RN [13]
 RP SUBCELLULAR LOCATION, AND MUTAGENESIS OF LEUCINE RESIDUES.
 RX MEDLINE=99415752; PubMed=10484781;
 RA Petris M.J., Mercer J.F.;
 RT "The Menkes protein (ATP7A; MNK) cycles via the plasma membrane both
 in basal and elevated extracellular copper using a C-terminal
 di-leucine endocytic signal.";
 RL Hum. Mol. Genet. 8:2107-2115(1999).
 RN [14]
 RP STRUCTURE BY NMR OF 375-446.


```

529 .....AGCATACAGGCGTCAACAATCATCAAGCCCTGCGTGGGC 573
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
193 leglylsleuglnlglyvalglnglylelyvalserleuaspnsngln 209
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
574 GAAGCAACCATGCTGCTGCGCCGACGACGCTCCCTCCCTCAGAGGCGG 623
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 GluAlaIThrIleValTyrGlnProHISLeuIleSerValGlnGluMetly 226
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
624 GGAAGCGGTATGGTGGATTTCTTCGCAACCTGCTAT 663
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
226 slvs.....GlnIleGluAlaMetClyPheProAlaPhe 237
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

seq_name: SwissProt_40:CYF_CHLVU

seq_documentation_block:
ID   CYF_CHLVU          STANDARD;          PRY;          341 AA.
AC   P56316;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Apocytochrome F precursor.
GN   PETA.
OS   Chlorella vulgaris.
OC   Chloroplast.
OC   Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC   Chlorellaceae; Chlorella.
NCBI_TaxID=3077;
RX   MEDLINE=97303241; PubMed=9159184;
RA   Wakasugi T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,
RA   Tsudzuki J., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A., Ohta T.,
RA   Inamuro A., Yoshinaga K., Sugita M.;
RT   *Complete nucleotide sequence of the chloroplast genome from the
RT   green alga Chlorella vulgaris: the existence of genes possibly
RT   involved in chloroplast division.
RL   Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).
CC   -1- FUNCTION: TRANSLOCATES PROTONS ACROSS THE THYLAKOID MEMBRANE AND
CC   TRANSFERS ELECTRONS FROM THE RIBSKE IRON-SULFUR PROTEIN AND PASSES
CC   THEM TO PLASTOCYANIN. THIS FUNCTION IS VERY SIMILAR TO THAT OF
CC   MITOCHONDRIAL CYTOCHROME C1.
CC   -1- SUBUNIT: MEMBER OF THE CYTOCHROME B6/F COMPLEX INCLUDING
CC   CYTOCHROME B6, CYTOCHROME F AND PROBABLY AN IRON SULFUR PROTEIN.
CC   -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane (Probable).
CC   -1- SIMILARITY: BELONGS TO THE CYTOCHROME C FAMILY.
CC   -----
CC   THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL: AB001684; BAA57987.1; -.
DR   HSP: P36438; IHC2.
DR   InterPro: IPR002325; Apocyt_F.
DR   InterPro: IPR000345; CytC_heme_bln.
DR   Pfam: PF01333; Apocytochrome_F; 1.
DR   PRINTS: PR00610; CYTOCHROME_F.
DR   PROSITE: PS00190; CYTOCHROME_C_1.
KW   Electron transport; Heme; Chloroplast; Thylakoid;
KW   Photosynthesis; Photosystem I; Photosystem II; Transit peptide;
KW   Transmembrane.
FT   TRANSIT 1..56 CHLOROPLAST (BY SIMILARITY).
FT   CHAIN 1..341 APOCYTOCHROME F.
FT   METAL 57..57 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   METAL 57..57 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   BINDING 77..77 HEME (COVALENT) (BY SIMILARITY).
FT   BINDING 80..80 HEME (COVALENT) (BY SIMILARITY).
FT   METAL 81..81 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT   TRANSMEM 307..327 POTENTIAL.

```

```

SQ SEQUENCE 341 AA; 37350 MW; FBC53FC5DE09465F CRC64;
alignment_scores:
  Quality: 100.50 Length: 312
  Ratio: 0.718 Gaps: 13
  Percent Similarity: 44.872 Percent Identity: 20.513
alignment_block:
US-09-303-518D-569 x CYF_CHLVU
Align seg 1/1 to: CYF_CHLVU from: 1 to: 341

```

```

7 CGTTTACATTCAGCGCTGTTTCCCTTTGGACCGCCATGCATCT 56
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6 LysleuGlnPheAsnPhelIleProAsnleuLys.....LysHISAlaVal 20
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
57 GTTGACCGCCCTGCTCAATGCTCTCCCTCCGCTTCCGCTGCTGCG 106
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
20 IPheserPheTrpGlyGlnAsnGlnAlaSnIleleuLysPheSerThrleu 37
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
107 ACAGCTGGGAACCGCGCTCGACATCTGCGTTTACCTTTAAAGGAA 156
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
37 alserLysGlyAlaIleuValIleuValCysSerPhePheleuThrAlaSer 53
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
157 GACCGCGCGCGCATCTC.....GCCATATGCGTCA 188
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 SerAsnAlaIatyrProIlePheAlaGlnGlnAsnTyrAlaAsnProAlaGly 70
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
189 GCGAGCGCATGANT..... 201
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
70 uAlaAsnGlyArgIleValCysAlaAsnCysHISleuAlaGluLysProI 87
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 ..... 201
87 legluIleGluValProGlnAlaValIleuProAsnPhrValPheGluAla 103
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
202 .....CCGACCCCAAAACGGCAAGCGCTTTTGGCGGAAC 239
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 ValValLysIleProIlyrAspLysGlnIleLysGlnValIleuAlaAsnG 120
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
240 GCGAAAGGC.....GGTT 253
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
120 YLysLysGlyAspLeuAsnValGlyAlaValIleuLeuProAspGlyP 137
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 TGGACCTGCGCCCGCGCTTTTACGAAACCGGAGACATAGAACAAAG 303
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
137 heGluIleAlaPro.....ProAspArgIleProGluGlu 148
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
304 TTCAAGCGGTACACGCGTGGGAACATGTGACAGCGCTTTGGACAAACA 353
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
149 MetLysAlaLysValGlyLysleuTyrPheGlnProTyrSerAlaGluLys 165
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
354 CGAAGGCTGCTATTCATCAGCGCGACATCGGACGCTTCGATTGGGCG 403
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
165 slvsThrIlePheValValGlyProValProGlyLys..... 177
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
404 GAGCGTACATCAGCAGCAGCTCCGCTCCGCTGACCGCATGTACAA 453
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
178 .....LysTyr.....SerGluMetValPheProIleleuSerPro..Asp 190
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
454 CCGCGCGCAATCAAGCATAGCAAAATC.....ATGCAAGCGG 494
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
191 ProAlaLysThrLysSerIleSerTyrleuLysTyrProIleTyrValG 207
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
495 CAGGGTTCGGCGCAAGAAAGAAACCGCGCTACACGATACATCAAGGGTCA 544
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 YGlyAsnArgGlyArgGlyGlnValTyrProAspGlySerLysSerAsn 224
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
545 AACCAATCAACAAAGCCCGCTGTCGCGAGCAACACATGCTCGGCC 594
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
224 snThrIlePheThrAlaSerAlaIleGlyLysIleThrAlaIle..... 238

```

```

595 GACACGTCCTCCCTCAAGAGGCGGAGGAGCGGTATGGTGATT 644
239 .....GluProAlaGlyLysGlyGly..... 246
645 CTTGGGCAACCTGCTATACCATGACGCTGGCGCAAAATTGGACAG 694
247 .....TyrThrLeuThrIle.....GluThrAlaAsnG 256
695 TCAAGGCGGTGAACCCCTGTTTTCGTCGCGACCGCTGCGCGGA 744
256 LysIleSerIleSer.....GluLysLeuProProGly 266
745 CAAGGTTTCATTTCACATCCCGCCGTCACAAAGG 780
267 ProGluLeuValValAsnIleGlyAspIleValGly 278

seq_name: Swissprot_40:DPD_ARATH

seq_documentation_block:
ID DPD_ARATH STANDARD: PRT: 1081 AA.
AC Q9LVN7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE DNA polymerase delta catalytic subunit (EC 2.7.7.7).
GN POLDI OR AT5G63960 OR MBM17.6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty.1 and TAC
RT clones."
RL DNA Res. 7:31-63(2000).
CC -1- FUNCTION: THIS POLYMERASE POSSESSES TWO ENZYMAITIC ACTIVITIES:
CC DNA SYNTHESIS (POLYMERASE) AND AN EXONUCLEOTIC ACTIVITY THAT
CC DEGRADES SINGLE STRANDED DNA IN THE 3' TO 5' DIRECTION.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate -> N diphosphate
CC + [DNA](n).
CC -1- SUBUNIT: HETERODIMER WITH SUBUNITS OF 125 kDa AND 50 kDa. THE 125
CC kDa SUBUNIT CONTAINS THE POLYMERASE ACTIVE SITE AND MOST LIKELY
CC THE ACTIVE SITE FOR THE 3'-5' EXONUCLEASE ACTIVITY (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC DIFFERENT REACTIONS OF DNA SYNTHESIS.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the European Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB019227; BAA96899.1; -
DR InterPro: IPR002064; DNA_pol_B.1.
DR Pfam: PF00136; DNA_pol_B.1.
DR PRINTS: PR00106; DNAPOLB.
DR SMART: SM00486; POLBC.1.
DR PROSITE: PS00116; DNA_POLYMERASE_B.1.
DR Transfaser: DNA-directed DNA polymerase; DNA replication;
DR DNA-binding; Hydrolase; Exonuclease; Zinc-finger; Nuclear protein.
FT ZN_FING 1005 1023 C4-type (POTENTIAL).

```

```

SQ SEQUENCE 1081 AA; 121735 MW; 152279F6291B06CF CRC64;

alignment_scores:
Quality: 100.00 Length: 225
Ratio: 0.862 Gaps: 12
Percent Similarity: 51.556 Percent Identity: 24.444

alignment_block:
us-09-303-518d-569 x DPD_ARATH ..

Align seg 1/1 to: DPD_ARATH from: 1 to: 1081

123 GCTCGACATCTGGCGTT.....TTACCTTTAAAGGACGCGCGGC 166
111 ||| |||..... ||| |||
404 AlaAlaThrLeuGlyIleGluGluPheProLeuGlyAlaGlyValAs 420
167 GCATGTCGCCCAATATGCGTCAGGCGCATGATCCGACCCCA.... 212
111 ||| |||..... ||| |||
420 nserArgValArgValArgAspSerThrPheSerSerArgInGlnGly 437
213 ..AACGGTCAAGCGCTTTTTCGGGAACGGCAAGCGCGTTGGAACT 260
111 ||| |||..... ||| |||
437 LeArgGluSerLysGluThrThrIleGluGlyArgPheGlnPheAspLeu 453
261 TCCCCCGCGTTTTCAGAAAACCGAAGACATAGAAACAATGTTCAAG 310
454 IlegAlaIle.....HisArgAspHisLysLeuSe 464
311 CGGTACACGCGTGGGACATGTGCAGACGCTTTGGACAAACAGAAAGG 360
111 ||| |||..... ||| |||
464 rSerArgValLeuAsnSerValSerAlaHisPheLeuSerGluInLysG 481
361 CTGCTATTTCATACCGCCGACATCGGACGATTCATTTGGCGGACGCTA 410
111 ||| |||..... ||| |||
481 LuAspValHisHisSerIleIleThrAspLeuGlnAsnGlyAsnAlaGl 497
411 CATGACGACGACGCTCCG...TTCCCGCTGACCGCCATGTACAAACGC 457
497 uThrArgArgArgLeuAlaValTyrCysLeuLysAspAlaTyrLeuPro 514
458 CGAAATCAAGCGATGACAAATCATG.....CAG 489
111 ||| |||..... ||| |||
514 InArgLeu.....LeuAspLysLeuMetPheIleTyrAsnTyrValGln 528
490 GCGGCGAGGTTCCGCGCAAGGAAAAACCGCGCTACACATA..... 534
529 MetAlaArgValThrGly.....ValProIleSerPheLeu 541
535 .....CAAGGGTCAACAAATCATCAAAAGCCCTCGTTGGGC. 573
111 ||| |||..... ||| |||
541 uAlaArgGlyGlnSerIleLeuValLeuSerGlnLeuAlaArgLysGlyL 558
574 ..GAGCAACCATGCTCTCCGACACGCTCCCTCCCTCAAGAGGC 621
111 ||| |||..... ||| |||
558 yslGlnLysAsnLeuValLeuProAsnAlaLysGlnSerGlyGln 574
622 GGG.....GAAGCGTATGGTC.....GATTCTTCGG 650
111 ||| |||..... ||| |||
575 GlYThrTyrGluGlyAlaThrValLeuGluAlaArgThrGlyPheTyr 591
651 CAACCTGCTATACCATGACGCTGGCGCAAAATGTGCACACGTCAAG 700
111 ||| |||..... ||| |||
591 uLysProIleAlaThrLeuAspPheAlaSerLeuTyrProSerIleMet 608
701 GCCTGAACACCTGTTTCTGCG 723
608 eValArgAsnLeuCysTyrCys 615

seq_name: Swissprot_40:AFSK_STRGR
seq_documentation_block:
ID AFSK_STRGR STANDARD: PRT: 807 AA.

```

AC P54742; 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Serine/threonine protein kinase afsk (EC 2.7.1.-).
 GN AFSK.
 OS Streptomyces griseus.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 NCBI_TaxID=1911;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96186909; PubMed=8635757;
 RA Ueda K., Uneyama T., Beppu T., Horinouchi S.;
 RT "The aerial mycelium-defective phenotype of *Streptomyces griseus*
 RT resulting from A-factor deficiency is suppressed by a Ser/Thr kinase
 RT of *S. coelicolor* A3(2).";
 RL Gene 169:91-95(1996).
 CC -1- FUNCTION: INVOLVED IN THE REGULATION OF SECONDARY METABOLISM BY
 CC PHOSPHORYLATING, ON BOTH SER AND THR, THE AFSK GLOBAL REGULATORY
 CC PROTEIN INVOLVED IN THE CONTROL OF SECONDARY METABOLISM (BY
 CC SIMILARITY).
 CC -1- PTM: AUTOPHOSPHORYLATED AT THR AND SER RESIDUES (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D45246; BAA08203.1;
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR Pfam: PF00069; Pkinase; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transferase; Serine/threonine-protein kinase; ATP-binding;
 KW phosphorylation.
 KM DOMAIN 16 272 PROTEIN KINASE.
 FT NP_BIND 22 30 ATP (BY SIMILARITY).
 FT BINDING 44 44 ATP (BY SIMILARITY).
 FT ACN_SITE 138 138 BY SIMILARITY.
 SQ SEQUENCE 807 AA; 85231 MW; 66C274219155D091 CRC64;
 alignment_scores: Quality: 98.50 Length: 220
 Ratio: 1.026 Gaps: 11
 Percent Similarity: 43.636 Percent Identity: 25.455
 alignment_block:
 US-09-303-518D-569 x AFSK_STRGR ..
 Align seg 1/1 to: AFSK_STRGR from: 1 to: 807
 82 TCCCTGCTGCGCTTCTGTCGACAGCTGGAAACCGGCTC..... 126
 190 SerValThhGlyAlaSerAspIlePheSerLeuGlySerThLeuValPh 206
 127GGACATCTGCGCTTTTACCTTTTAAAGAAAGACCGCGGC 166
 206 eAlaIaThhGlyHisAlaProPheHis..... 215
 167 GCATGCTGCCAATATGCTGACGCGCATATCCCAACCCCAAAAG 216
 216GlyAlaAsnPro..... 219
 217 GTCAAGCGCTTTTTCGGAACGCGAAGCGGCTTGGAACTTGCCTCC 266
 |||:|||||:|||||: |||

220 ValGIuThrValIhneMetLeuValArgGIuGly.....Pr 231
 267 CGCGTTTTCAGAAACCGGAGACATAGAAACATGTTCAAGCGGTAC 316
 231 AspLeuIuGlyLeuProAspAspLeuArgProLeuIleGlySerGysm 248
 317 ACGGCTGGGACATGTGCGACGCGCTTGGACAAACAGAGGCGTCTGA 366
 248 etGIuMetAspAlaThhHisProAlaIaGIuProArgAspLeuGlnAla 264
 367 TTCATCAGCGCGGCATC.....GGCAGTACGATTTGGCGGAGC 407
 265 GlnLeuAlaProHisLeuPheAlaSerGIySerAspSerGIy..... 279
 408 CTACATCGCCAGCAGCTCCGCTCCGCTGACCGCCATGTACAACGCGC 457
 280 ThraIaSerAlaIaThhProValProAlaThraIaMetIleGIuArg 295
 458 CGAAATCAAGCGCATAGCAAAATCATGACGCGGCGGCGTTCGCGC 507
 296 ArgArgGIyGIyArgArgThr.....Al 303
 508 AAGGAAAAACCGCGCTACGACGATACAGGCGTCAACAAATCATGAA 557
 303 aArgArgProProArg..... 308
 558 AGCCCGCTTGGGCGAAGCAACCATGCTCCTGCCGACGCTCCCT 607
 309Pro.ArgProAlaGIyLeuArgAla 317
 608 CCCCTCAAGAGGCGGAGGCGATGATGGTGATTTCTTCGCAAACT 657
 317 IaProGIuGIyProGIyAlaGIyHisArgLeuAlaGIuArgGIyArgPro 333
 658 GCCTATACCATGACGCTGCGCGCAAAATGGCACACGTCAAAGCGGTGA 707
 334 AlaPheAlaLeuProAlaValLeuAlaIaValaIaArgValArg 350
 708 AACC 711
 350 gThr 351
 seq_name: Swissprot_40:AT7A-CRIGR
 seq_documentation_block: PRT; 1476 AA.
 ID AT7A-CRIGR STANDARD; PRT; 1476 AA.
 AC P49015; 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Copper-transporting ATPase 1 (EC 3.6.3.4) (Copper pump 1)
 DE (Fragment).
 GN ATP7A.
 OS Cricetulus griseus (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Cricetulus.
 NCBI_TaxID=10029;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-K1;
 RX MEDLINE=96154683; PubMed=8589689;
 RA Camkarakis J., Petris M.J., Bailey L., Shen P., Lockhart P.,
 RA Glover T.W., Barcroft C., Patton J., Mercer J.F.;
 RT "gene amplification of the Menkes (Mnk; ATP7A) P-type ATPase gene of
 RT CHO cells is associated with copper resistance and enhanced copper
 RT efflux";
 RL Hum. Mol. Genet. 4:2117-2123(1995).
 CC -1- FUNCTION: MAY FUNCTION IN THE EXPORT OF COPPER FROM THE CYTOPLASM
 CC TO AN INTRACELLULAR ORGANELLE. IT MAY SERVE AS WELL FOR THE EXPORT
 CC OF OTHER METALS.
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + Cu(2+)(IN) = ADP + PHOSPHATE +
 CC Cu(2+)(OUT).

```
alignment_scores:
  Quality: 98.50
  Ratio: 0.794
  Percent Similarity: 53.680
  Length: 231
  Caps: 14
  Percent Identity: 25.108
```

Align seg 1/1 to: AT7A_CRIGR from: 1 to: 1476

[illegible]

179 ATATCCGTCAGCGACGACGTAAATCCGACCCCAAAAACGGTCAAAGCCGTT 228
::: :
53 sPrOlylsLeu.....GlntHrProLystHLeugIngluaL 65
229 TTTGGCGAAMCGGCMAAACGGCGATTG....GAACCTGCCCGCGGTT 272
::: :
66 IleAspArMetGIyRheAspAlaLeuEnlIsAsnIaAsnProLeu.. 81
273 TTTCAGAAAAACCGCAGACACATAGACAACATGTTCAAGCGGTACAGCG. 321
|||
82ProValLeuthrAspTrHleupRheLeuthrValThralas 95
322TGGAACCATGTGGACAGCGCTTTGGACAAACGGA 357
||:||||:||||:||||:||||:||||:||||:||||:||||:
95 eHeuThrLeuProTfprAspHisIleGlnSerThrLeuEuls...Thry 111
358 GGGC.....TGTATTTCATCACGCGCCACATCGGCGACTA 392
||| |
111 sgLYvalThrAspRIeLySIleRheProGInLSarGLthrLeuAlaValt 128
393 CGATTGTGGCGGAGCGTACATGACGACGAGCGTTCGGTCCCGGTGACCG 442
::: ||:||||:||||:||||:||||:||||:||||:||||:
128 hrIleIleProSerIleValIsAsnIaAsnInlleuysGIueuValPro 144
443 CCATGTTCAA.....ACCGCCGMAAATCAACG..... 470
::::::::::||| |::::::::::|||
145 .GIueuSerLeuGIuThrgLythrLeuGIuLSarSergLyIaCySg 161
::::::::::||| |::::::::::|||
471 ..GATACACAAATCATGCGAGCGGGC.....AGGTTCCGGG 506
||||:||||:||||:||||:||||:||||:||||:||||:
161 IuAspHis..SerMetIaGInIaLGluValIValLeuLySIleYsVa 177
507 CAAGAAGAAACCgcCGcCTAcC..... 528
177 lGIuGIuMetHrChySHisSerCySHrSerThrHrthruGIuLySIleG 194
::: ||| |::: |||
529 ..AGCATACAAAGGGTCAACAACATCATCAAAAGCCGTTCGGCGGAA 576
::: ||:||||:||||:||||:||||:||||:||||:||||:
194 LyusLeuGIuNGluValGIuArgIleLyValSerLeuAspIndInLIu 210
577 GCACCATGTCCTCGCCGACCAACGCGCCCTCCCTCAAAAGCGCGGGA 626
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
211 AlAtHrIleValLyGIuInProHisIleuIeserValGIuGIuIleLySy 227
627 AGCGGTATGGGTGATTTCTTCGCGCAACCTGCGCTAT 663
::: ||:||||:||||:||||:||||:||||:||||:||||:
227GInIleGIuAlaMeGIyPheProAlaPhe 237

seq_name: SwissProt_40:RPOC_CHLTR

seq_documentation_block:

ID RPOC_CHLTR STANDARD; PRT; 1396 AA.
OS 084316;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DR 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase
beta' chain) (RNA polymerase beta' subunit).
GN RPOC OR CT314.
OC Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RX NCBI_TaxID=813;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/W-3/CX;
RA MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatsov R.L., Zhao Q., Koornin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RT Science 282:754-759(1998).
LC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION

CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
 CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
 CC BETA' CHAIN.
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE001304; AAC67907.1;
 CC InterPro: IPR000722; RNA_pol_A.
 CC Pfam: PF00623; RNA_pol_A; 1.
 CC Transferrase: DNA-directed RNA polymerase; Transcription;
 CC Complete proteome.
 CC KW
 CC SEQUENCE 1396 AA; 154904 MW; B24BF841D284065 CRC64;

alignment_scores: 98.00 Length: 332
 Quality: 0.662 Gaps: 15
 Ratio: 0.662 Percent Identity: 20.181
 Percent Similarity: 44.578

alignment_block:

US-09-303-518D-569 x RPOC_CHLTR ..

Align seg 1/1 to: RPOC_CHLTR from: 1 to: 1396

51 CATCTGTTGACCGCCCTGCTCAATGCGCTCTCCCTGCGCGTTTCT 100
 |||||
 1084 HSPRO.....GlnllealalleTyralaaspalaasme 1095
 101 GTCTGCACACCGTGGGAACCGCGCTCGACATCTCGCGTTTACTTTA 150
 :::: |||||
 1095 TlysgluLeuValGlyThrTyralalleProserelYalaalle.Ileer 1111
 151 AAGGAAGACCGCGCGCATGCTGCGCAATATGCGTCAAGCGCATGAA 200
 |||||
 1112 ValGluGluGlnGlnArgllealaleProglymetleuValaArgleu.. 1127
 201 TCCGACCCCAAAACGTCAAAGCCGTTTTCGGAACGCGCAAAAGCG 250
 |||||
 1128ProArg.....GlyAlaalleTyThrLysaspilleThrglyG 1140
 251 GTTTG.....GAACCTGCCCGCGGTTTTCAGAAACCGGAA 288
 |||||
 1140 TlyleuProArgValaIaGluLeuValaGluA.....ArgLysProGlu 1154
 289 GACATAGAAACATG.....TTCAAGCGGT 314
 |||||
 1155 AspAlaIaaspillealaleLysleaspglyValaIaaspheLysglYl 1171
 315 ACACGGCTGGGAACATGTCACAGCGCTTGGAC.....A 349
 :::: |||||
 1171 eGlnLysasnLysArglleLeuValaIaArgspelleThrglyMetG 1188
 350 AACACGAGGCTGCTATTCATCAGCGCGCATC.....GGC 387
 :::: |||||
 1188 IuGluIuHisLeuIleSerleuThrLysHisLeuIleValaGlnArgly 1204
 388 AGCTACGATTTGGCGGACGCTACATCAGCAGCAGCTTCCCTCCGCT 437
 :::: |||||
 1205 AspSerValIleLysGlyGlnGlnLeuThrAspGlyLeuValaIaProH 1221
 438 GACCGCGATG.....TACAAACCGCGGAATCAAGGATAG 475
 :::: |||||

1221 sGluIleuGlnIleCysGlyValaArgGluLeuGlnLysTyLeuVala 1238
 476 ACAAAATCATCAGCGCGCGGCTGCGGC.....AAGGAAA 516
 :::: |||||
 1238 sngluValGlnGluValaTyArgLeuGlnGlyValaIaaspheLysAspLys 1254
 517 ACCGCGCTACAGCATCAAGGGGTCAACAAATCATCAAGCCCTGCG 566
 :::: |||||
 1255 HsValGluIlelleValaArgGlnMetleuGlnLysValaArgIleThas 1271
 567 TTCGCGGACAGCAACCATCTGCTGCGCGACAGCTCCCTCC..... 609
 :::: |||||
 1271 perGlyAspThrThrLeuLeuPheGlyGluAspValaAspLysGluP 1288
 610CCTCAGAAGCGCGGAGCGGTATGGGTG 639
 :::: |||||
 1288 heTyGluGlnAsnArgArgThrGlnGluAspGly..... 1299
 640 GATTTCTTGGCAACCTGCTATTCATCAAGCTG..... 675
 :::: |||||
 1300GlyLysProAlaGlnAlaValaProValleuGlyIleTh 1313
 675 675
 1313 rLysAlaSerleuGlyThrGluSerPheIleSerAlaIaSerPheGlna 1330
 676GCGGCAAAATGGCACAC 693
 :::: |||||
 1330 spThrThrArgValleuThraAspAlaIaCysSerSerLysThrAspThr 1346
 694 GTCAAGCGCTGAAAACCTGTTTCTGCTGCGACCGCTGCGCG 743
 :::: |||||
 1347 LeuGlyGlyPheLysGlnAsnValIleMetGlyIleMetIleProGlyG 1363
 744 ACNAGGTTTCAATTGACATCCGCCCGCTCCAGGGAATTGAAGCGG 793
 :::: |||||
 1363 yThrGlyPheAspThrHisLysArgIleLysGln.....HisLeuG 1377
 794 ACAAGCCCATGATGCGCGCTGTTTCAACCGCAATGCCAATAT 837
 :::: |||||
 1377 IuLysGluGlnIuAspLeuValaIaPheAspPheAspSerIuPhe 1391
 seq_name: SwissProt_40:ALR_STRCO
 seq_documentation_block:
 ID ALR_STRCO STANDARD: PRT: 391 AA.
 AC 086786;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Alanine racemase (EC 5.1.1.1).
 GN ALR OR SC664.23.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=A3(2);
 RC Saunders D., Harris D., Parkhill J., Barrett B.G., Rajandream M.A.;
 RL Submitted (Aug-1998) to the EMBL/Genbank/DBSJ databases.
 CC -1- CATALYTIC ACTIVITY: L-alanine -> D-alanine.
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
 CC -1- SIMILARITY: BELONGS TO THE ALANINE RACEMASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

DR EMBL: AL031317; CAA20401.1; -
 DR HSSP: P10724; 1BD0.
 DR InterPro: IPR000821; Ala_raceemse.
 DR Pfam: PF00842; Ala_raceemse; 2.
 DR PRINTS: PR00992; ALARACEEMASE.
 DR PROSITE: PS00395; ALANINE_RACEEMASE; 1.
 DR Isomerase; Pyridoxal phosphate.
 KW BINDING 46 46
 FT PYRIDOXAL PHOSPHATE (BY SIMILARITY).
 SQ SEQUENCE 391 AA; 41199 MW; 24754AC1385DCAIB CRC64;

alignment_scores:

Quality: 97.50 Length: 190
 Ratio: 1.096 Gaps: 12
 Percent Similarity: 46.842 Percent Identity: 29.474

alignment_block:

US-09-303-518D-569 x ALR_STRCO ..

Align seg 1/1 to: ALR_STRCO from: 1 to: 391

```

246 AGGCGGTTGGAACTGCCCCCGCTTTTCAGAAA.....ACGGGAG 289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
130 ArgAlaAlaIaIyAlProAlaArgValGlnLeuLysAlaAspThrGlyLe 146
290 ACATAGAACATGTTCAAAAGCGGTACAGCGGCGGAGACATGTCAGCAG 339
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
146 uGlyArgGlyGlyCysGlnProGlyAla AspTrpLeuArgLeuValGly 162
340 GCTTTG.....GACAAACAGAGGCGTCTATTCATCAGC..... 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
163 AlaAlaLeuArgAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 179
376 .CGGACATCGGCGAGCTAGCATTTGGCGCGGAGCTACATCAGCCAGCAG 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
179 pSerHisPheAlaCysAlaAspGlnProGlyHis..... 190
425 TTCGGTCCCGCTGACCGCATGTACAAACCGCGAAATCAAGCGAG 474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
191 .....ProSerIleAlaAla... 195
475 GACAAATATGTCAGCGCGGCGAGGCTTCGCGCAAAAGAAACCGCGCC 524
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
196 .....GlnLeuThrArgPheArg.....GlnMetThrAlaTyr 206
525 TACGACATACAGAGGCTCAAAACATCATCAAGCCCTCGCTGGGCGC 574
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
206 rAlaGlnGlnArgGlyLeuArgProGlnValArgHisIleAlaAsnSerP 223
575 AAGCAACATGCTCCGCCGCGAC...CACGTC..... 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
223 rGlaTThrLeuThrLeuProAspAlaHisPheAspLeuValArgProGly 239
604 .....CCTCCCTCCAGAGCGCGGAGGAGCGGT 632
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
240 IleAlaMetTyrGlyValSerProSerProGlnIleGlyThrPro.... 254
633 ATGGGTGATTTCTTCGCGCAAACTGCTATACATGACGCTGGCGCGCA 682
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
255 ....AlaAspPheGlyLeuArgPro.....ValMetThrLeuAlaHis 268
683 AATTGGCAGCAGTCAGGCGGTGAAGAACCCGTTTCTTCTGCGGAACGC 732
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
268 erLeuAlaLeuValLys.....Gln 274
733 CTGCCTGGCGAGCAAGGT 750
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
275 ValProGlyGlyHisGly 280

```

seq_name: SwissProt_40:NONA_DROME

seq_documentation_block:

ID NONA_DROME STANDARD; PRT; 700 AA.

AC Q04047;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE No-on-transient A protein.
 GN NONA.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxId=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90262721; PubMed=2344408;
 RA Jones K.R., Rubin G.M.;
 RT "Molecular analysis of no-on-transient A, a gene required for normal
 vision in Drosophila.";
 RL Neuron 4:711-723(1990).
 CC -1- FUNCTION: REQUIRED FOR NORMAL VISION IN DROSOPHILA.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; I (SHOWN HERE) AND II; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: M33496; AAA03214.1; -
 DR EMBL: M33496; AAA03215.1; -
 DR HSSP: P19339; 1SXL.
 DR Flybase: FBgn0004227; nona.
 DR InterPro: IPR000504; RRM.
 DR Pfam: PF00076; rrm; 2.
 DR SMART: SM00360; RRM; 2.
 DR PROSITE: PS0102; RRM; 2.
 DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
 KW RNA-binding; Repeat; Vision; Alternative splicing.
 FT DOMAIN 302 374
 FT DOMAIN 376 457
 FT DOMAIN 73 76
 FT DOMAIN 254 264
 FT DOMAIN 619 626
 FT DOMAIN 648 652
 FT VARSPLIC 666 700
 DSFAFEFGVNNMGCGNCGNNGGNNVPMGRRRF -> VC
 FHPKRYPTKYSVTNRSVELQELLSMIPMKL (IN
 ISOFORM II).
 SQ SEQUENCE 700 AA; 76967 MW; 732DB77FC5DFD47 CRC64;

alignment_scores:

Quality: 97.50 Length: 297
 Ratio: 0.733 Gaps: 11
 Percent Similarity: 44.781 Percent Identity: 21.212

alignment_block:

US-09-303-518D-569 x NONA_DROME ..

Align seg 1/1 to: NONA_DROME from: 1 to: 700

```

153 GGAAGACCGCGCGCATGTCGCCAATAT.....GCTCAGG 190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
44 GlyGlyProAlaGlnLysLysGlnArgPheGlyGlyProAsnAlaGlnAs 60
191 CAGGCATGATCCCGACCCCAAAAGCT..... 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
60 ngInAsngInAsngInAsngInAsngInAsngInAsngInAsngInAsng 77
219 .....CAAACCGTTTTCGCGAAGCGCAAAAGCGG 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

77 1aValGlyGlyProAsnGlnAsnLysAsnPhcGlyAsnAsnLysGlyGly 93
252 TTTTGGAACTGGCCCCGCTTTTTCAGAAAACCGGAGACATAGAAACA 301
94 Phe.....ValGlyAsnArgAsnArgAsnAs 102
302 TGTTCAAAGCGGTACACG..... 320
102 naaAArgAlaGlyAsnGlnAsnArgThrPheProGlyAsnAsnAsnSerA 119
320 ..... 320
119 snGlnLysProAsnAsnGlnThrSerLysAlaAspGlyProAsnAlaLeu 135
321 .....CTGGACATGTGCAGACAGCTTTGGACAAACAG 355
136 AlAlAsnAsnAsnGlnProAlaThrAlaAlaAlaGlyGlnAsnGlnAlaAs 152
356 AAGGCGTGTCTATTCATCAGCCGCCACATCGGACGTAAGATTGGGGCGA 405
152 ngLAsnAlaAsnLysGlyLAsnGlnArgGln.....GlyGlnA 166
406 CGCTACATCAGCCAGACCTTCCGCTCCGCTGACCCGCTGACAAACC 455
166 snGlnAsnGlnAsnGlnVal.....HisGlyGlnGly 176
456 GCCGAAATTCAAAGCGATACAAATCATGCAGCGCGGAGCGGCTTGGC 505
177 AsnGlnGlyGlyProGlyAsnGlnGlyAlaGlyAsnGlnGlyGly 193
506 GCAAGGAAAGAACCGCGCTACACATACAAAGGGGTCAA..... 545
193 ngLAsnGlnGlyGlyAlaGlyAsnGlnGlyAsnGlyGlnGlyPheArg 210
546 .....ACAAATCATCAAGCCCTGGCTCGG..... 572
210 LArgAsnAlaGlyAsnAsnGlnGlyGlyPheSerGlyGlyProGln 226
573 .....CGAGCAACCATGCTGCTGCGGACCGACCTCCCTCCCTC 613
227 AsnGlnGlnArgAsnArgAsnArgSerGlyProArgPro..... 240
614 AAGAAGGCGGAGGCGTATGGGTGATTTCTTGGCAACCTGCGCAT 663
241 .....GlyGlyGlyAlaGlyAlaMetAsnSerThrAsnMetG 254
664 ACCATGAGCGCGCGGCAAAATTTGSCACACGTCAAAGCGTGAACCCCT 713
254 LylGlyGlyGlyGlyGlyGlyGlyGlyProArgGlyGlyGlnAsp 270
714 GTTTTCTGCTGCGAAGCCTGCTGCGGAGCAAGGTTTGCATTTGCACA 763
271 PheHeIethrGlnArgLysArgSerIleSerGlyProThrPheGln 287
764 TCCGCCCGCTCCAA..... 777
287 euGlnProValGlyValProThrGlnThrLysPheSerGlyArgAsnArg 303
778 .....GGGGAATTGAACGCGCAAAAGCCCATGAT 807
304 LeuTyrValGlyLysLeuThrAsnAspIleThrAspAsp 316

seq_name: SwissProt_40:5E5_RAT

seq_documentation_block:
ID 5E5_RAT STANDARD; PRT; 825 AA.
AC 063003:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 5E5 antigen.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; Tissue=Brain;
RX MEDLINE=96015159; PubMed=8537300;
RA Suzuki E., Kojima N., Yoshimura K., Oyemura K., Obata K., Akagawa K.;
RT Cloning and sequence analysis of cDNA for a possible DNA-binding
RL protein 5E5 in the nervous system.;
RU J. Biochem. 118:122-128(1995)
CC -1- FUNCTION: MIGHT HAVE DNA-BINDING ABILITY.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN NEURONS.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: D37934; BAA07153.1;
KW DNA-binding; Nuclear protein; Antigen.
SQ SEQUENCE 825 AA; 86831 MW; AF667FE2FD55BDF CRC64;

```

```

alignment_scores:
Quality: 97.50 Length: 294
Ratio: 0.780 Gaps: 12
Percent Similarity: 42.517 Percent Identity: 22.449

```

```

alignment_block:
US-09-303-518D-569 x 5E5_RAT ..

```

```

Align seg 1/1 to: 5E5_RAT from: 1 to: 825

```

```

6 TCGTTTCAATTCAGCGCTGTTCCTCCCTTTGGCAACCGCATGACATCC 55
535 SerGlnThrLeuProAlaLeuAlaGlyAlaProThrAlaHisAla 551
56 TGTTCAGCGCGCTGTCAAATGCGCTGCTGCGGCTTTCCTGTCG 105
551 aValProGlyProGly.....ProAlaAlaIleThrLeuGlyG 564
106 CACACGCTGGGAACCGGCTGCGACATCTGGCGTTTACCTTTAAAGA 155
564 Ly..... 564
156 AGACCGCGCGCATGCTGCGCAATGTCGTCAGGACGATGATCCG 205
565 .....ArgLysArgThrGlySerThrPargLysGlyArg..... 575
206 ACCCCAAACGTCAAAGCGCTTTTGGGAAAGCGCAAAAGCGGTTG 255
576 .....ArgGlyGlyGlyAlaGlyAlaSerGlyGlyGlyArgGly 591
256 GAACCTGCCCCCGGCTTTTCAGAAAACCGGAGACATGAAACAATGTT 305
591 Ly.....ArgGlyArgGlyLysArgGlyGly 601
306 CAAAGCGGTACACG.....CTGGACATGTGCAGACAGCTTTGGACA 349
602 LeuSerGlyThrArgGlyAspAlaGlySerProSerIleArgArgGly 618
350 AACACGAAGGCTGCTATTCATACCGCGCAATCGCACGCTACGATTGG 399
618 uGlnArgArgArgGly..... 623
400 GCGGAGCGTATCATCAGCAGACCTTCCGCTGACCGGCATGTA 449
624 .....HisGlyProProAlaAlaGlyAlaAlaGlnValSerThr 636

```



```

450 CAAMCCCGCAAAATCAAGAGATAGACAAATCATGACGGCGGAGG 499
    : : : : : : : : : : : : : : : : : : : : : : : :
637 ArgGlyArgGlyAlaArgGlyGlnArgThrGlyGlnAlaGlnAspG 653
    : : : : : : : : : : : : : : : : : : : : : : : :
500 TTCCGGCGCAAGAAACCGCGCTACGACATACAGGCGGTCAACAA 549
    : : : : : : : : : : : : : : : : : : : : : : : :
653 yLeuLeuProArgGlyArgGlyAspArgLeuProLeuArgProGly 670
    : : : : : : : : : : : : : : : : : : : : : : : :
550 ATCAT.....CAAGCCCTGCTCGGCGCAAGCAACCAT..... 584
    : : : : : : : : : : : : : : : : : : : : : : : :
670 sngInArgValGlnArgProGlyHisProArgGlyGlnGlyAlaLe 686
    : : : : : : : : : : : : : : : : : : : : : : : :
585 .....CCTCTGCGCGCAACGCTCCCT 607
    : : : : : : : : : : : : : : : : : : : : : : : :
687 AsnAlaProSerAlaProAspAlaSerProGlnHisProArgArg 703
    : : : : : : : : : : : : : : : : : : : : : : : :
608 CCCCTCAAGAGCGGCGGAGAG.....CCTATGGGTGCA. 641
    : : : : : : : : : : : : : : : : : : : : : : : :
703 pValSerGlnGlnArgGlnArgLeuThrArgGlnPheArgValG 720
    : : : : : : : : : : : : : : : : : : : : : : : :
642 .....TTT 644
    : : : : : : : : : : : : : : : : : : : : : : : :
720 lypHepProProProProThrArgProProValLeuLeuProLeu 736
    : : : : : : : : : : : : : : : : : : : : : : : :
645 CTTGGGCAACCTGCTATACCATGACGCTGCGGCAAAATGGACAGC 694
    : : : : : : : : : : : : : : : : : : : : : : : :
737 LeuArgLeuThrCysAlaGlyAspProGlyAlaSerArgProGly 753
    : : : : : : : : : : : : : : : : : : : : : : : :
695 TCAAGAGCGTGAAACCTGTTTCTGCTGCAAGCGCTGCGGCGGA 744
    : : : : : : : : : : : : : : : : : : : : : : : :
753 gArg.....ProAlaArgArgP 759
    : : : : : : : : : : : : : : : : : : : : : : : :
745 CAAGGTTCGATTGCATCCGCCCGTCCA 776
    : : : : : : : : : : : : : : : : : : : : : : : :
759 rokArgGlyLeuLeuThr...ProGlnArgPro 768
    : : : : : : : : : : : : : : : : : : : : : : : :
seq_name: SwissProt_40:RPOC_CHLMU
seq_documentation_block:
ID RPOC_CHLMU STANDARD; PRT; 1396 AA.
AC 09PK79;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase
DE beta' chain) (RNA polymerase beta' subunit).
GN RPOC OR TC0588
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / N199;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouli H., Craven B., Bowman C., Dodson R.,
RA Gelin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC - FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC - CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC [RNA] (N).
CC - SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN.
CC - SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002327; AAF39420.1; -
DR TIGR; TC0588; -
DR InterPro; IPR000722; RNA_POL_A.
DR Pfam; PF00623; RNA_POL_A; 1.
KM Transferase; DNA-directed RNA polymerase; Transcription;
KW Complete proteome.
SQ SEQUENCE 1396 AA; 154897 MW; CC090B8A18868977 CXC64.

alignment_scores:
    Quality: 97.00      Length: 332
    Ratio: 0.660      Gaps: 15
    Percent Similarity: 44.277      Percent Identity: 20.181

alignment_block:
US-09-303-518D-569 x RPOC_CHLMU ..
Align seg 1/1 to: RPOC_CHLMU from: 1 to: 1396

51 CATCTGTTGACCGCCCTGCTCAATGCTTCCTCCCTGCGGCTTTCCT 100
    : : : : : : : : : : : : : : : : : : : : : : : :
1084 HisPro.....GlnLeuAlaLeuArgAlaAspAlaAsnMe 1095
    : : : : : : : : : : : : : : : : : : : : : : : :
101 GTCGTGACAGCGTGGGAAACCGCTCGACATCTGCGCTTTACCTTTTA 150
    : : : : : : : : : : : : : : : : : : : : : : : :
1095 tGlnGlnLeuValGlyThrTYrAlaAlaIleProSerGlyAlaIle. Ileser 1111
    : : : : : : : : : : : : : : : : : : : : : : : :
151 AAGGAAGACCGCGCGCGCATGCTGCGCAATATGCGTCAAGCAGCGCATGAA 200
    : : : : : : : : : : : : : : : : : : : : : : : :
1112 ValGlnGlnGlyGlnArgIleAlaProGlyMetLeuAlaArgLeu.. 1127
    : : : : : : : : : : : : : : : : : : : : : : : :
201 TCCCGACCCCAAAACGTCGAAGCCGTTTGGCGAAACGGCAAAAGCG 250
    : : : : : : : : : : : : : : : : : : : : : : : :
1128 .....ProArg.....GlyAlaIleLysThrLysAspIleThrGlyG 1140
    : : : : : : : : : : : : : : : : : : : : : : : :
251 GTTTC.....GACCTGCGCGCGGCTTTTCAGAAACCGCGAA 288
    : : : : : : : : : : : : : : : : : : : : : : : :
1140 lYLeuProArgValAlaGlnLeuValGlnAla.....ArgLysProGlu 1154
    : : : : : : : : : : : : : : : : : : : : : : : :
289 CACATAGAAACAAATG.....TTCAAGCGGT 314
    : : : : : : : : : : : : : : : : : : : : : : : :
1155 AspAlaAlaAspIleAlaLysIleAspGlyValValAlaAspHelysGly 1171
    : : : : : : : : : : : : : : : : : : : : : : : :
315 ACACGGCTGGGACATGTGCGACGAGGCTTTGGAC.....A 349
    : : : : : : : : : : : : : : : : : : : : : : : :
1171 eGlnLysAsnLysArgIleLeuValValArgAspGlnValThrGlyMetG 1188
    : : : : : : : : : : : : : : : : : : : : : : : :
350 AACCAAGAGGCTCTATTCATACGCGCGACATC.....GGC 387
    : : : : : : : : : : : : : : : : : : : : : : : :
1188 lGlnGlnLysLeuIleSerLeuThrLysIleuLeuValGlnArgGly 1204
    : : : : : : : : : : : : : : : : : : : : : : : :
388 AGTACAGTTCGGGCGGACGCTACATCAGCCACGACTTCGCTCCGCT 437
    : : : : : : : : : : : : : : : : : : : : : : : :
1205 AspSerValIleLysGlyGlnGlnLeuThrAspGlyLeuValValProH 1221
    : : : : : : : : : : : : : : : : : : : : : : : :
438 GACCGCATG.....TACAACCGCGCAAAATCAAGCATG 475
    : : : : : : : : : : : : : : : : : : : : : : : :
1221 sGlnIleLeuGlnIleCysGlyValArgGlnLeuGlnLysTYrLeuValA 1238
    : : : : : : : : : : : : : : : : : : : : : : : :
476 ACAAAATCATGACGCGGCGAGGTTCCGCGC.....AAGAGAAA 516
    : : : : : : : : : : : : : : : : : : : : : : : :
1238 sngInValGlnGlnValTYrArgLeuGlnGlyValAspIleAsnAspLys 1254
    : : : : : : : : : : : : : : : : : : : : : : : :
517 ACCGCGCTTACAGCATACAGGCGTCAACAAATCATGACCCCTCGG 566
    : : : : : : : : : : : : : : : : : : : : : : : :
1255 HisIleGlnIleLeuValArgIleMetLeuGlnLysValArgIleThrAs 1271

```



```

1271 yscInLeuAlaSerHisGlnAlaValHisMetGluArgArgGlyGly 1287
348 CAACAGAGAGGCTGCTATTCATCAGCCGACATCGCAGTACGATT 397
1288 GLyThrArgLysAlaThrArgGluAspArgProPheArgGlyGly 1304
398 TGGCGGAGCGTACATCAGCAGCAGCTTCGTCGCGGACCGGACG 447
1304 sGlyArgThrTyr.....ArgHisA 1311
448 TACAACCGCCGAAATCAAGCAGTACAAATATCATCAGCGGCGAG 497
1311 laGlySerLeuLeuAsnHis.....ArgArgSerHisGlnThrGlyGln 1325
498 GGTTCGCGGCAAGGAAAAACCGGCTACCGACATACAGAGGCGTCAAC 547
1326 TyrSerCysProThrCysProLysThrTyrSerAsnArgMetAlaLeu 1342
548 AAATCATCAAGCCCTGCGTTCGCGGCAAGCAACATCGTCTCGCCGAC 597
1342 sAspHisGln.....ArgLeuHisSerGluAsnAlaGlyGly 1354
598 CAGGTCCCTCCCTCAAGAGCGGCGGAGCGCTATGGGT..... 638
1354 rGlyArgAlaGlySerArgArgThrAlaValArgCysAlaLeuGly 1370
639 .....GGATTTCCTCGCAACCTCGCTATACCATGA 670
1371 ArgSerPheProGlyLysGlySerLeuGluArgHisLeuAlaGlyHisGln 1387
671 C.....GCTGCGCGCAAAATTTGGCACAC 693
1387 uGlnThrGluArgGluProAlaAsnGlyGlnGlyLeuAspGlyThrA 1404
694 GTCAAGCGCTGAAACCTGTTTCTGCTGCGCAAGCCGCTGGCGGG 743
1404 lAlaSerGlnAlaAsn.....LeuThrGlySerGlnGlyLeuGlu 1417
744 ACAAGTTTCGATTTCACATCCG.....CC 769
1418 ThrGlnLeuGlyGlyAlaGluProValProHisLeuGluAspGlyValPr 1434
1434 oArgProGly.....GluArgSerGlnSerPro 1443
seq_name: SwissProt_40:NTC1_HUMAN
seq_documentation_block:
ID NTC1_HUMAN STANDARD; PRT; 2444 AA.
AC P46531;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurogenic locus notch protein homolog 1 precursor (Translocation-
DE associated notch protein TAN-1) (Fragment).
GN NOTCH1 OR TAN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91347367; PubMed=181692;
RA Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,
RA Smith S.D., Sklar J.;
RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by
RT chromosomal translocations in T lymphoblastic neoplasms.";
RL Cell 66:649-661(1991).
CC -!- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCTE FUNCTION. IN
CC ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION
CC IN SOME T-CELL NEOPLASMS.

```

```

CC CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC CC -!- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN,
CC CC BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT
CC CC IS FOUND MAINLY IN LYMPHOID TISSUES.
CC CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
CC CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC CC -!- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC CC -----
CC CC between the Swiss-Prot entry is copyright. It is produced through a collaboration
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR DR EMBL: M73980; AAA60614.1; -.
DR DR HSPSP: P00740; IIXA.
DR DR MIM: 190198; -.
DR DR InterPro: IPR002110; ANK.
DR DR InterPro: IPR000152; Asx_hydroxyl.
DR DR InterPro: IPR000561; EGF-like.
DR DR InterPro: IPR000742; EGF_2.
DR DR InterPro: IPR001881; EGF_Ca.
DR DR InterPro: IPR000800; Notch.
DR DR Pfam: PF00023; ank; 6.
DR DR Pfam: PF00008; EGF; 36.
DR DR Pfam: PF00066; notch; 3.
DR DR SMART: SM00248; ANK; 5.
DR DR SMART: SM00179; EGF_CA; 22.
DR DR SMART: SM00001; EGF_Like; 13.
DR DR SMART: SM00004; NL; 2.
DR DR PROSITE: PS50088; ANK_REPEAT; 4.
DR DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR DR PROSITE: PS00010; ASX_HYDROXYL; 20.
DR DR PROSITE: PS00022; EGF_1; 34.
DR DR PROSITE: PS01186; EGF_2; 26.
DR DR PROSITE: PS01187; EGF_CA; 18.
DR DR Differentiation: Neurogenesis; Repeat; ANK repeat; EGF-like domain;
DR DR Transmembrane; Signal; Glycoprotein.
DR DR SIGNAL 1 18
DR DR CHAIN 19 >2444
DR DR TRANSMEM 19 1736
DR DR TRANSMEM 1737 1757
DR DR DOMAIN 1758 >2444
DR DR DOMAIN 20 58
DR DR DOMAIN 59 99
DR DR DOMAIN 102 139
DR DR DOMAIN 140 176
DR DR DOMAIN 178 216
DR DR DOMAIN 218 255
DR DR DOMAIN 257 293
DR DR DOMAIN 295 333
DR DR DOMAIN 335 371
DR DR DOMAIN 372 410
DR DR DOMAIN 412 450
DR DR DOMAIN 452 488
DR DR DOMAIN 490 526
DR DR DOMAIN 528 564
DR DR DOMAIN 566 601
DR DR DOMAIN 603 639
DR DR DOMAIN 641 676
DR DR DOMAIN 678 714
DR DR DOMAIN 716 751
DR DR DOMAIN 753 789
DR DR DOMAIN 791 827
DR DR DOMAIN 829 868
DR DR DOMAIN 870 906
DR DR DOMAIN 908 944
DR DR DOMAIN 946 982
DR DR DOMAIN 984 1020
DR DR DOMAIN 1022 1058

```

FT	DISULFID	645	655	BY SIMILARITY.
FT	DISULFID	650	664	BY SIMILARITY.
FT	DISULFID	666	675	BY SIMILARITY.
FT	DISULFID	682	693	BY SIMILARITY.
FT	DISULFID	687	702	BY SIMILARITY.
FT	DISULFID	704	713	BY SIMILARITY.
FT	DISULFID	720	730	BY SIMILARITY.
FT	DISULFID	725	739	BY SIMILARITY.
FT	DISULFID	741	750	BY SIMILARITY.
FT	DISULFID	757	768	BY SIMILARITY.
FT	DISULFID	762	777	BY SIMILARITY.
FT	DISULFID	779	788	BY SIMILARITY.
FT	DISULFID	795	806	BY SIMILARITY.
FT	DISULFID	800	815	BY SIMILARITY.
FT	DISULFID	817	826	BY SIMILARITY.
FT	DISULFID	833	844	BY SIMILARITY.
FT	DISULFID	838	855	BY SIMILARITY.
FT	DISULFID	857	867	BY SIMILARITY.
FT	DISULFID	874	885	BY SIMILARITY.
FT	DISULFID	879	894	BY SIMILARITY.
FT	DISULFID	896	905	BY SIMILARITY.
FT	DISULFID	912	923	BY SIMILARITY.
FT	DISULFID	917	932	BY SIMILARITY.
FT	DISULFID	934	943	BY SIMILARITY.
FT	DISULFID	988	999	BY SIMILARITY.
FT	DISULFID	993	1008	BY SIMILARITY.
FT	DISULFID	1010	1019	BY SIMILARITY.
FT	DISULFID	1026	1037	BY SIMILARITY.
FT	DISULFID	1031	1046	BY SIMILARITY.
FT	DISULFID	1048	1057	BY SIMILARITY.
FT	DISULFID	1064	1075	BY SIMILARITY.

alignment_scores:	quality:	97.00	Length:	275
Ratio:	0.815		Gaps:	19
Percent Similarity:	43.273		Percent Identity:	23.636
alignment_block:				
US-09-303-518D-569 x NTCL_HUMAN ..				
Align seg 1/1 to: NTCL_HUMAN from: 1 to: 2444				
2	TGTTTGCTTACAAATTCAGCGCTGTTCCCTTTCGCAAGCCCA.....	46		
89	CysAlaLeuGlyPheSerGly.....ProLeuGlyLeuThrProLeuAs	103		
47TGCACATCCTGTTGACCGCCCTGTCAATACCTTCCTCCG.	88		
103	pasAlaLeuGlyLeuThr.....AsnProGlyArgAsnGlyLeuThrCysA	118		
89TGGCGCTTCCGTGTGCAACAGC	112		
118	spleuLeuThrLeuThrGluTyrLysCysArgCysPro..ProGlyTyrPse	134		
113	TGGGA.....CCGCTCGGA	129		
134	rgLYLysSerCysGlnGlnAlaAspProCysAlaSerAsnProCysAlaA	151		
130	CATGTGC.....GTTTACCTTTAAAGAAAGCCGCGCGCATCGC	176		
151	snGlyGlyGlnCysLeuProPheGlnAlaSerTyrIleCysHisCysPro	167		
177	CAATATGCGTAGGC.....AGCATGAATCCCGACCCCAAAAGG	217		
168	ProSerPheHisGlyProThrCysArgGlnAspValAsnGlnCysGlyI	184		
218	TCAAGCCGTTTTCGGAAAGCGGAAAGGCGGTTTGGACATTCGCCCC	267		
184	nlySerThrArgLeuCysArgHisGly.....GlyThrCys..	195		
268	GCGTTTTCAGAAACCGGACAGACATAGAACATGTTCAAACGGTACA	317		
196HisAsnGlnValGlySerTyrArg..CysValCysArgAlaThrH	210		

```

318 CGGCTGGAGACATGTGACAGAGCTTGGACAAACAGAGGCTGCTAT 367
   ||||| ||||| ||||| |||||
210 s.ThcglyProAsnGlyValProTyrVal.....ProCysSer 223
368 TCATCAGCCCGCACATCGGAGCTACGATTTGGCGGAGCTACATCAGC 417
   |||
224 ProSer..... 225
418 CAGCAGCTTCGCTCCGCTGACCGCATATCAAAACCGCGAAATCAAA 467
   ||||| |||
226 .....ProCysGlnAsn..... 229
468 AGCCATGACAAAATCATGACGAGCGGGGAGGTTCCGGCGAAAGAAAA 517
   ||||| ||||| ||||| |||||
230 .....GlyGlyThrCysArgProThrClyAspValThrHisGlyCys 243
518 CCGCGCTACCGCATATCAAGGGGTCAAAACAAATCATCAAGCCCTGCGT 567
   ||||| ||||| ||||| |||||
244 AlacysLeuProGlyPheThrGlyGlnAsn.....Cys.. 254
568 TCGGGCGAAGCAACCATCGTCTGCGCGACGACGCTCCCTCCCTCAGA 617
   ||||| ||||| ||||| |||||
255 .....GlnGlnAsnIleAspAspCysProGlyAsnAsn.....CysLysA 268
618 AGGGGGGAGAGGGGTATGGGTGATTTCTTGGCAAAACCTGCTATACCA 667
   ||||| ||||| ||||| |||||
268 snelgylalacysValAspGlyValAsnThrTyAsnGlyProCysPro 284
668 TGAGCTGGCGGCAAAATTTGCCACAGCTCAAGGGGTGAAGACCTGTTT 717
   ||||| ||||| ||||| |||||
285 ProGluThrClyThrClyThrCysThrGluAspVal.....AspGlyCysG1 300
718 TTCTGCTGGGAAGCCTGC 736
300 nleuMetProAsnAlaCys 306

seq_name: SwissProt_40:VE2_HPV38

seq_documentation_block:
ID VE2_HPV38 STANDARD; PRT; 441 AA.
AC 080910:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Regulatory protein E2.
GN E2.
OS Human papillomavirus type 38.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=37959;
RN [1]
RA Delius H.;
RT Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACNNNNNGGT-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC REPLICATION.
CC -!- SUBUNIT: BINDS DNA AS A DIMER.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL: U31787; AAA9453.1; -
DR HSP: P17383; 1DHM.
DR InterPro: IPR001866; E2_N.
DR InterPro: IPR000427; Early2_C.
DR Pfam: PF00511; E2_C; 1.
DR Pfam: PF00508; E2_N; 1.
DR ProDom: PD000672; Early2_C; 1.
DR ProDom: PD000678; E2_N; 1.
DR KW Early protein; transcription regulation; Activator; DNA-binding;
KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ SEQUENCE 441 AA; 49731 MW; A820DC2CD05ED601 CRC64;

```

```

alignment_scores:
Quality: 96.50 Length: 357
Ratio: 0.656 Gaps: 18
Percent Similarity: 41.176 Percent Identity: 23.529

```

alignment_block:

US-09-303-518D-569 x VE2_HPV38

```

Align seg 1/1 to: VE2_HPV38 from: 1 to: 441
18 CAGGCTGTTCCCTTTCGCAACCGCCATGCAATCTGTTGACCGCC 67
   ||| :::: ||||| |||||
40 GlnIleIleThrHisThrAlaArgThrHisGly.....ValThrArgLe 54
68 TGCTCAAAATGCTCTCCCTGCTGCGCGCTTCTCTGTCACACGCTGGGA 117
   :::: ||||| ||||| |||||
54 uGlyTyrGlnProValProSerLeuAla.....SerSerGluAlaLysA 69
118 AACCGCTCGGACATCTGGCGTTTACCTTTTAAG..... 153
69 lAlaYsPaLaIleSerMetValIleu.LeuLeuGlnSerLeuLysLysSe 85
154 .....GAAGACGGCGCGCATGCTGCGCAATATGCGTCAGCAG 193
85 rLysTyrAlaAspGlnGlnIleThrPheLeuAlaGlnThrSerLeuGluAla 102
194 GCATGAAATCC.....GACCCCAAAACGGTCAAA 222
102 aLArgSerProAlaAspCysPheLysLysGlyProLysAsnIleGlu 118
222 GCGGTTTT.....GCGGAAACGGCAAAAGCGGTTTGAACCTGTC 263
::||| |||||
119 ValValIlePheAspGlyAspProGluAsnLeuSerTyrThrValTrpTh 135
264 CCGCGCGTTTTCAGAAACCGGAAGACATGAAACAAATGTCAAAGCGG 313
::||| |||||
135 rTyrlleTyrTyrLeuThrAspGluAspIle..... 145
314 TACAGCGCTGGGAACATGTGACGAGCGCTTGGAC..... 348
||||| |||||
146 .....TrpGluLysValGlnGluYHisValAspTyrThrGlyAlaTyr 159
349 AAACGCAAGGCGCTGATTCATC..... 372
::||| |||||
160 TyrTyrGlnGlyLysLeuLysValTyrTyrLeuLysPheGluAsnAspAl 176
372 ..... 372
176 aLysArgTyrGlyValThrGlyLeuThrPheGluValHisValAsnLysAspT 193
373 .....ACCGCGACATCGGC.....AGCTAC 393
193 hValPheThrProValThrSerSerThrProValGlyLysSerThr 209
394 GATTGGGCGGAGCTA.....CAVCAAGCAGAGCTTCGTTCCCGC 436
||| ::||| |||||
210 AspSerAlaSerArgAlaAlaLeuProGluProSerThrInSerValSerP 226
437 TGACCGC...CATGTCAACACCGCGCAAAATCAAGACATGACAAATTC 483

```


281 rGAspLeuSerLeuSerGluAlaArgIleProValProLysGluVal... 296
 607 TCCTCTCAGAGCGGCGGAGGTATGGGTATTTCTTCCGCAACC 656
 297ArgGluGlyGlyGlyVal...GluArgPheProGlyGlyP 310
 657 T.....GCTTACCATGACCTGCGCGCAAAATGG 688
 310 CTYrArGArGTyrPheAlaAlaLeuTyrArgAlaLeuGlyGluAla 327
 689 CACAGCTAAAGCGGTGAAACCTGTTTCTGCTCCGACCCCTGCT 738
 327 euserThnGlyLeuAlaArgAlaLeuValAlaGlyLysGlyLeu 343
 739 GCGGACAGGTTTC.....GATTTCACATCCGCCCTGCCA 776
 344 GluGlyValAlaGlyLeuAlaGlyValAlaGlyAlaPheLeuArgProLeuG 360
 777 AGGGGANTTGAC 789
 360 uAlaArgLeuSer 364

seq_name: SwissProt_40:H1LR_CHICK

seq_documentation_block:
 ID H1LR_CHICK STANDARD; PRT; 218 AA.
 AC P08288;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Histone H1.11R.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87250632; PubMed-3597432;
 RA Coles L.S., Robins A.J., Madley L.K., Wells J.R.E.;
 RT "Characterization of the chicken histone H1 gene complement.
 Generation of a complete set of vertebrate H1 protein sequences.";
 RL J. Biol. Chem. 262:9656-9663(1987).
 CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
 NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M17020; AAA48790.1; -;
 DR PIR: C28456; C28456.
 DR HSP: P08287; IGHC.
 DR InterPro: IPR001386; Linker_histone.
 DR Pfam: PF00538; Linker_histone; 1.
 DR SMART: SM00526; H15; 1.
 KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
 FT INIT_MET 0
 FT DOMAIN 37 110 GLOBULAR.
 SO SQUENCE 218 AA; 21672 MW; CB9724BFF14654A6 CRC64;

alignment_scores:
 Quality: 95.00 Length: 272
 Ratio: 0.709 Gaps: 15
 Percent Similarity: 49.265 Percent Identity: 25.735

alignment_block:
 US-09-303-518D-569 x H1LR_CHICK ..
 Align seg 1/7 to: H1LR_CHICK from: 1 to: 218

54 CCTGTGACCGCCCTGCTCAAAATGCTTCCTCCGCTGCTTCTCTGTC 103
 5 ProAlaAlaAlaProAlaAlaAlaProAlaProAlaAla.....Ly 18
 104 TGCACAGCTGTGGAACCGGCTGGACATCTGGCGTTTACCTTTAAAG 153
 18 sAlaAlaAlaLysLysProLysLysAlaAlaGly.....G 30
 154 GAAGACCGCGCGCATCGTCCCAATATGCGTCAGACGATGATGCC 203
 30 LysAlaLysAlaArg.....LysProAlaGlyProSerVa 41
 204 CGACCCCAAAACGTCAAACCTTTTTCGGCAACGCAAAAGCGGTT 253
 41 LThrGluLeuIleThrLysAlaValSerLysGluArgLysGlyL 58
 254 TCGAATCTTGCCCGCGGTTTTCAGAAACCGAAGACATAGAACATG 303
 58 euserLeuAla.....AlaLeu 63
 304 TTCAAACGCTACACGCTGGACATG.....TGCAGCA 338
 64 LysLysAlaLeuAlaAlaGlyGlyTyrAspValGluLysAsnAsnSer 80
 339 GCGTTTGACAAACACGAGGCGTCTATCATCGCCGCGCATCGGCA 386
 80 rGlyLeuLysGlyLeuLysSerLeuValSer.....LysIly 92
 389 GCTACATTTGGGCGAGCGTACATCAGCCAGC...ACCTTCGTTCCG 435
 93 ThrLeuValGlnThrLysGlyThrGlyAlaSerGlySerPheArgLeu 109
 436 CTGACCGCATGTACAAACCGCGCAAAATCAAGAGATAGCAAAATCAT 485
 109 rLysLysProGlyGlyGlyLeuGlyLysAlaProLys..... 121
 486 GCAGCGGCGGAGGTTTCGCGCAAGAAACCGCGCTCAGCATAC 535
 122 ..LysLysAlaSerAlaAla.....LysProLysLysAlaAla 134
 536 AAGGGTCAAAACAATCATCAAGCCCTGCGTGGGGAACAACCATC 585
 135 Lys.....LysProAlaAlaAlaAlaLysLysProly 145
 586 GTCTGCGCCGACACGCTCCCTCCCTCAAGAGCGGGAAGCCGATG 635
 145 sLysAlaValAlaValLysLysSerProLysLysAlaLysProAla 162
 636 GGTGATTTCTTGGCAACCTGCTATACCATGACCTGGCGCAAAAT 685
 162 LaSerAlaThrLysSerValLysSerPro..... 172
 686 TGGCAGACGTCAAAGCGCTGAACCTGTTTCTGCTCGGAACGCTG 735
 173LysLysAlaAlaLysPro..... 178
 736 CCGCGGCAAGGTTTCATTTGCATCCGCGCGCTCCAGGGAAT 785
 179LysLysAlaValAlaAlaLysSerProAlaLysAlaLys. 191
 786 GAAGCGGCAAAACCATGATGCGCGCGTTCACACGCAATGCCGAT 835
 192AlaValLysProLysAlaAlaLysProLysAlaLysProLys 206
 836 ATTGATACGCC 847
 207 AlaAlaLysAla 210

RL Biochemistry 32:11345-11351(1993).
 CC - FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
 CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
 CC - SUBCELLULAR LOCATION: Nuclear.
 CC - SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M17019; AAA48789.1; -
 DR PIR: B28456; B28456.
 DR PDB: 1GHC; 31-AUG-94.
 DR InterPro: IPR001386; Linker_histone.
 DR Pfam: PF00538; Linker_histone; 1.
 DR SMART: SM00526; H15; 1.
 KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
 KW Acetylation; 3D-structure.
 FT INIT_MER 0 0
 FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
 FT DOMAIN 40 113 GLOBULAR.
 SQ SEQUENCE 224 AA; 22397 MW; D3D057CB97865CAF CRC64.

alignment_scores:
 Quality: 95.00 Length: 267
 Ratio: 0.714 Gaps: 14
 Percent Similarity: 49.813 Percent Identity: 25.843

alignment_block:

US-09-303-518d-569 x H1L1.CHICK

Align seg 1/1 to: H1L1.CHICK from: 1 to: 224

```

78 CCTCTCCGCTGGCGCTTCCGCTGTCACACGCTGGGAAACCGGCTCG 127
   ||| ||||| ||| ||| ||||| ||||| |||||
5 PROIALPROIALAIAIAAIAAIAAIAAIAAIAAIAAIAAIAAIAAIA 21
128 GACATCTGGCGCTTTTACCTTTAA.....GGAAGACGCGCGCGC 168
   : : : : : ||| ||| ||| : : : : :
21 SALAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIA 37
169 ATGCTGCCAATATGCTGACGACGACGATCCGACCCCAAAACGAT 218
   : : : : : ||| ||| : : : : :
38 .....LysProAlaGlyProSerValThGluLeuIleTh 49
219 CAAGCGCTTTTGGGAAACGCAAGCGCTTTGCACTTGGCCCG 268
   : : : : : ||| ||| : : : : :
49 RYSAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIA 64
269 CGTTTTCAGAAACCGGAAGACATAGAAACAAATGTTCAAGCGGTAC 318
   : : : : : ||| ||| : : : : :
65 .....AlaLeuLysLysAlaLeuAl 71
319 GGCTGGGAACATG.....TCGACGAGCTTTGGACAAACA 353
   : : : : : ||| ||| : : : : :
71 AIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIAIA 88
354 CGAAGCGCTGATATCATACGACGCGCAATATGATATTTGGCG 403
   ||| : : : : : ||| : : : : :
88 eulysSerLeuValSer.....LysGlyThLeuValGlnThr 100
404 GACGCTATACGACGACG...AGCTTCGCTTCGCGTACGCGCATGAC 450
   : : : : : ||| ||| : : : : :
101 LysGlyThGlyAlaSerGlySerPheArgLeuSerLysProGly 117
451 AAACGCGCAAAATCAAGCATATAGCAAAATCATGACGCGCGAGGT 500
   : : : : : ||| ||| : : : : :
117 uValLysLysLysAlaProLys.....LysLysAlaSerA 129

```

seq_name: SwissProt_40:VE2_HPV05

seq_documentation_block:

ID VE2_HPV05 STANDARD; PRT; 514 AA.

```

AC P06921;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Regulatory protein E2.
GN E2.
OS Human papillomavirus type 5.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10578;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87207670; PubMed=3033892;
RA Zachow K.R., Ostrow R.S., Farris A.J.;
RT "Nucleotide sequence and genome organization of human papillomavirus
RT type 5."
RL Virology 158:251-254(1987).
CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCNNNNGGT-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC REPLICATION.
CC -1- SUBUNIT: BINDS DNA AS A DIMER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M17463; AAA46986.1; -
DR PIR: D26277; W2MT5.
DR HSSP: P03122; 2BOP.
DR InterPro: IPR001866; E2_N.

```



```

335  gaArgProGlnValProGly.....AlaArgProArgProSerPro. 348
223  GCGGTTTTTCGGGAACGGCAAAAGCGGTTTGGAACTTCCCGCGGCTT 272
349  .....GlyThrValPro..... 352
273  TTTGAGAAACCGGAGACATAGAAACAATGTTCAAAGCGGTACAGCGCT 322
353  ....SerGlnProValArgHisArgGlnCysAlaGlnGlyValThrArgSe 368
323  GGGAAACATGTGCAGCAGCGCTTGGACAAACACGAGCGGTGCTATTTCATC 372
368  rAspArgArgAlaGlyGlyGlyArg..... 377
373  ACSCCGCACATCGGAGCTACGATTGGGGGCGAGCTACATCAGCCAGCA 422
377  ..... 377
423  GCTTCGCTCCGCTGACGCCCATGTACAAACCGCAAAATCAAGCA 472
378  GlySerGlyProAlaGly.....AspAlaAspCysArgSerAs 390
473  TAGACAAATTCATGCAGCGGCGGAGGTTCCGCGCAAGAAACCCGCG 522
390  pProPArgGlnArgAlaAlaGlyArgGlyGlnAlaAspTrpGlyAspA 407
523  CCTACCGCATCAAGGGGTCAACAAATCAACAAACCCGCTGTCGGG 572
407  La.....HisLeuArgArgGlnAla.....AlaGlyArg 416
573  CGAAGCAACCATCTCTCCGCCACACGCTCCCTCCCTCAAGAAGCG 622
417  ArgAlaGlnHisTrpSerLeuGlnPro.....AlaAlaAspAlaGlyArg 432
623  GGAAGAGCGTATGGGATTTCTTCGGCAACCTGCTATACCATGACG 672
432  yAlaLeuArgAlaAspAlaCysLeuArgGlnGlyCysProGlnHisHis 449
673  CTGGC.....GGCAAAATTCGGCAGCAGTCACAAAG 701
449  IsGlyCysHisHisLeuSerAlaArgArgAspLeuGlyGlnArgGlnArg 465
702  CGTGAACCCCTGTTTCTGCTGGGAACGCTGCTGCGGACAAAGTT 751
466  .....LeuArgProGlyAla..GlyProAspLeuG 475
752  TCGATTTC.....CACATCCGCGCCGTCGAAGGGAATGACGGCGAC 795
475  LuAspLeuCysArgHisAlaGlyGlnGlnGlyAlaGlyGlyAsp 491
796  AAGGCCCATGATCCCGCGCTTCAACCGCAATGCGCAATTTGAGTACG 845
492  ArgSerSerAlaAla.....ArgArgAla.....TrpLeuArg 502
846  CCGTTT.....CCGACGACAGT 862
502  gLeuSerGlySerMetAspTrpValThrThrSerAspAspProThrPro 519
863  ATCTG 867
519  yTrLeu 520
seq_name: SwissProt_40:SON_HUMAN
seq_documentation_block:
ID SON_HUMAN STANDARD; PRT; 2426 AA.
AC P18583; Q95981; Q9UPY0; Q14120; Q14487; Q9UPR9; Q9H7BL; Q9P070;
AC Q9P072;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE SON protein (SON3) (Negative regulatory element-binding protein) (NRE-

```

```

DE binding protein) (DBP-5) (Bax antagonist selected in saccharomyces 1)
DE (BASS1) (protein C21orf50).
GN SON OR NREBP OR DBP5 OR C21ORF50 OR KIAA1019.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM A; B; C; D; E AND F).
RX MEDLINE=21564202; PubMed=11707072;
RA Raymond A., Friedl M., Neergaard, Henrichsen C., Chapot F.,
RA Deutsch S., Ucla C., Rossier C., Lyle R., Guipponi M.,
RA Antonarakis S.E.;
RT "From PREDs and open reading frames to cDNA Isolation: revisiting the
RT Human Chromosome 21 Transcription Map.";
RT Genomics 78:46-54(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM G).
RC TISSUE=Liver;
RX MEDLINE=21316479; PubMed=11306577;
RA Sun C.-T., Lo W.-Y., Wang I.-H., Lo Y.-H., Shiou S.-R., Lai C.-K.,
RA Ting L.-P.;
RT "Transcription repression of human hepatitis B virus genes by negative
RT regulatory element-binding protein/SON.";
RT J. Biol. Chem. 276:24059-24067(2001).
RN [3]
RP SEQUENCE OF 1-689 FROM N.A. (ISOFORM H).
RC TISSUE=Placenta;
RA Casadel R., Strippoli P., D'Addabbo P., Canaleir S., Lenzi L.,
RA Vitale L., Giannone S., Carinci P., Zannotti M.;
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-130 FROM N.A.
RC TISSUE=Smooth muscle;
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senda T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y., Ota T., Suzuki Y.,
RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA Isogai T., Sugano S.;
RT "NEO human cDNA sequencing project.";
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-114 FROM N.A.
RC TISSUE=Blood;
RA Ye M., Zhang O.H., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
RA Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
RT Human partial CDS from cd34+ stem cells.";
RT Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 437-2426 FROM N.A. (ISOFORM B).
RC TISSUE=Brain;
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirosewa M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RT DNA Res. 6:197-205(1999).
RN [7]
RP SEQUENCE OF 554-2426 FROM N.A. (ISOFORM A).
RX MEDLINE=92049296; PubMed=1944255;
RA Chumakov I.M., Berdichevskii F.B., Sokolova N.V., Reznikov M.V.,
RA Prasolov V.S.;
RT "Identification of a protein product of a novel human gene SON and
RT into mammalian cells.";
RT Mol. Biol. (Mosk) 25:731-740(1991).
RN [8]
RP SEQUENCE OF 709-1079 FROM N.A. (ISOFORM I).
RC TISSUE=Placenta;
RX MEDLINE=93062885; PubMed=1435774;
RA Bliskovskii V.V., Kirillov A.V., Zakhariev V.M., Chumakov I.M.;
RT "The human son gene: the large and small transcripts contains various

```



```

272 TTTTCAGAAACCGGAGACATAGAAACATGTTCAAAACGGCTACAGCC 321
      :::::::::: :::::::::::::::::::: ::::: |||
1892 ...LysArgSerProLysHisArgSerLysArgGluArgLysArgL 1907
      ::::: ::::: :::::::::::::::::::: :::::
322 TGGACACATGTCGACAGGCTTTGGACAAACAGCAGGCTGATTCAT 371
      ::::: ::::: :::::::::::::::::::: :::::
1907 ysaArgSerSerArgAspAsnArgLysThrValArgAlaArgSerArg 1923
      :::::::::::::::::::: :::::
372 CAGCCGACATGGCAGCTACGATTTGGCGGAGCTACATACGCCAGC 421
      :::::::::::::::::::: :::::
1924 ThrProSerArgArgSerArgSerHisThrProSerArgArgArgSe 1940
      :::::::::::::::::::: :::::
422 AGCTTCGGTTCCGCTGACCGCCATGTACAAACCGCCGAAATCAAGC 471
      :::::::::::::::::::: :::::
1940 rArgSerValGlyArgArgArgSerPheSerIleSerProSerArgArg 1957
      :::::::::::::::::::: :::::
472 ATGACAAATTCATGACGCGGCGGAGGCTTCGCGCAAGAAACCGC 521
      :::::::::::::::::::: :::::
1957 eraArg.....ThrProSerArgArgSerArg 1965
      :::::::::::::::::::: :::::
522 GCCTACAGCATACAAAGGGGTCAAAACATCATCAAGCCCTGCTCG 571
      :::::::::::::::::::: :::::
1966 ThrProSerArgArgSerArgThrProSerArgArgSerArgThrPro 1982
      :::::::::::::::::::: :::::
572 GCGACGACATCATGCTCTGCTCCGACACCTCCCTCCCTCAAGAGC 621
      :::::::::::::::::::: :::::
1982 rArgArgSerArgThrProSerArgArgSerArgThrProSerArgArg 1999
      :::::::::::::::::::: :::::
622 GCGGAAGGCGTATGGGTGATTTCTTCGCAAAACGCTATACCATGAC 671
      :::::::::::::::::::: :::::
1999 rArgArgSerArg..... 2002
      :::::::::::::::::::: :::::
672 GCTGCGGCGCAAAATTCGACACGCTCAAGAGCGTGAACCGCTTTTCT 721
      :::::::::::::::::::: :::::
2003 .....SerValValArgArgArgSerPheSerIleSerProValArg 2017
      :::::::::::::::::::: :::::
722 G.....CTGCGAAGCGCTGCGCTGCGGCGCAAGGTTTCATTTGCACATC 765
      :::::::::::::::::::: :::::
2017 ValArgArgSerArgThrPro.....LeuArgArgArgPheSerArgS 2031
      :::::::::::::::::::: :::::
766 CGGCC.....CGTCCAGAGGGAATTAACGCGGCAACAGCCCA 803
      :::::::::::::::::::: :::::
2031 eProlleArgArgLysArgSerArgSerSerGluArgGlyArgSerPro 2047
      :::::::::::::::::::: :::::

```

seq_name: SwissProt_40:SGS3_DROER

seq_documentation_block:

ID SGS3_DROER STANDARD; PRT; 328 AA.

AC P13730;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Salivary glue protein Sgs-3 precursor.

GN SGS3.

OS Drosophila erecta (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7220;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=88332966; PubMed=3138416;

RA Martin C.H., Mayeda C.A., Meyerowitz E.M.;

RT "Evolution and expression of the Sgs-3 glue gene of Drosophila.";

RL J. Mol. Biol. 201:273-287(1988)

CC -1- DEVELOPMENTAL STAGE: PRODUCED BY THIRD-INSTAR LARVAE.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL, M14652; -; NOT_ANNOTATED_CDS.

DR PIR: S01359; S01359.

DR FLYBASE: FBgn0012266; Dere\SGS3.

KW Repeat; Signal.

FT SIGNAL 1 23

FT CHAIN 24 328 SALIVARY GLUE PROTEIN SGS-3.

SQ SEQUENCE 328 AA; 36355 MW; 62F27F18C0DF8272 CRC64;

alignment_scores:

Quality: 94.50 Length: 231

Ratio: 0.836 Gaps: 14

Percent Similarity: 48.918 Percent Identity: 27.706

alignment_block:

US-09-303-518D-569 x SGS3_DROER ..

Align seq 1/1 to: SGS3_DROER from: 1 to: 328

```

2 TGTTCGTTTACATTCAGCGCTGTTCCCGCT.....TGCGA 39
  ||| ::::: ||| ||:::
37 CysThrThrAlaArgProThrCysAlaProValThrThrThrCysAr 53
  ::::: ::::: :::::
40 ACCGCGATGACATCC...TGTGACGCGCTGCTCAATGCTCCCT 86
  ::::: ::::: ||| ||:::
53 gProThrThrThrThrArgCysProProThrThrThrArgCysProp 70
  ::::: ::::: ||| ||:::
87 GTCGCGCTTCCTGTC...TGCACAGCGTGGGAACCGCTCGACATC 133
  ||::: ||::: ||| |||
70 rOrProThrArgProAlaGluCysThrAla...ThrThrLysArgProThr 85
  ::::: ::::: ||| |||
134 TGGCGTTTACCTTTTAAAGAGAGACGCGCGCATGTCGCCAATATG 183
  ||| ||| ||| ||:::
86 AlaArgProThrThrArgArgThrThrValArgAlaThrThr..... 99
  ::::: ::::: ||| |||
184 GTCGAGGAGGATGATCCCGACCCCAAAACGCTCAAGCCGTTTTCG 233
  ::::: ::::: ||| |||
100 ...LysArgAla.....ThrThrArgArgThrThrLysArgAla 112
  ::::: ::::: ||| |||
234 GAAAGCGCAAAAGCGGTTTGAACCTTGCCCGCGCTTTTTCGAAAC 263
  ||::: ||::: ||| |||
112 hrThrArgArgThrThrVal..... 118
  ::::: ::::: ||| |||
284 CGGAAGACATGAAACATGTTCAAGCGGTACAGCGCTGGCAATGTG 333
  ||| ||| ||| ||:::
119 ArgAlaThrThrThrLysArgAlaThrThrArgArgThrThrThrLysArgAl 135
  ::::: ::::: ||| |||
334 CAGCAGGCTTTGACAAACAGAGGCGCTGCTATTC.....T 371
  ::::: ::::: ||| |||
135 aProThrArgArgThrThrThrLysArgAlaThrThrArgArgAsnPro 152
  ::::: ::::: ||| |||
372 CAGCGCGCATCG.....GCAGTACGATTTGGGCGGAGCGTAC 412
  ::::: ::::: ||| |||
152 hrArgArgThrThrThrArgArgAlaProThrLysArgAlaThrThrLys 168
  ::::: ::::: ||| |||
413 TCAGCCGACGCTTCGTTCCCGCTGACCGCCCATGTACAAACCGCCGAA 462
  ||::: ||::: ||| |||
169 ArgAlaThrThrThrArgAsnProThrLysArgLysThrThrArgArgThr 185
  ::::: ::::: ||| |||
463 ATCAAGCGATGACAAATCATGACGCGCGGAGGTTGCGGCAAGG 512
  ::::: ::::: ||| |||
185 rThrValArgAlaThrThrThrLysArgAla.....ThrThrLysA 200
  ::::: ::::: ||| |||
513 AAAAACG.....CGCTACGACATACAGGG 541
  ::::: ::::: ||| |||
200 rGlaProThrLysArgAlaThrThrLysArgAlaProThrLysArgVal 216
  ::::: ::::: ||| |||
542 TCAAAACATCATCAAGCCCTGCGTTGGGG.....AACCAAC 582
  ::::: ::::: ||| |||
217 ThrThrLysArgAla...ProThrLysArgAlaThrThrLysArgAla 232
  ::::: ::::: ||| |||

```

583 ATGCTCTGCGCCGACGACGTTC...CTCCCTCAAGAGCG 622
 |||||
 232 CTHTLysArgAlaThrLysArgAlaProThrLysArgAla 246

seq_name: SwissProt_40:SOHC_ZYMMO

seq_documentation_block: PRT; 658 AA.
 ID SOHC_ZYMMO STANDARD;
 AC P3390;
 DT 01-FEB-1994 (rel. 28, Created)
 DT 01-FEB-1994 (rel. 28, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Squalene-hopene cyclase (EC 5.4.99.-).
 GN SHC.
 OS Zymomonas mobilis.
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
 OC Zymomonas.
 OX NCBI_TaxID=542;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 31821 / ZM4 / CP4;
 RX MEDLINE=95202074; PubMed=7894707;
 RA Reipen I.G., Poralla K., Sahm H., Sprenger G.A.;
 RT "Zymomonas mobilis squalene-hopene cyclase gene (shc): cloning, DNA
 sequence analysis, and expression in Escherichia coli."; Microbiology 141:155-161(1995).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-ATCC 31821 / ZM4 / CP4;
 RA Reipen I.G., Sahm H., Sprenger G.A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 31821 / ZM4 / CP4;
 RA Um H.W.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CATALYZES THE CYCLIZATION OF SQUALENE INTO HOPENE.
 CC -1- PATHWAY: KEY ENZYME IN HOPANOID (TRITERPENOID) METABOLISM.
 CC -1- SUBCELLULAR LOCATION: Membrane-associated (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE TERPENE CYCLASE/MUTASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 4 PEST REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X80766; CAA56749.1; -;
 DR EMBL: X73561; CAA51958.1; -;
 DR EMBL: AJ001401; CAA04735.1; -;
 DR EMBL: AF203881; AAF12829.1; -;
 DR PIR: S37494; S37494.
 DR HSSP: P33247; 15QC.
 DR InterPro: IPR001330; Prenyltrans.
 DR InterPro: IPR002365; terpene_synth.
 DR Pfam: PF00432; prenyltrns; 4
 DR PROSITE: PS01074; TERPENE_SYNTHASES; 1.
 DR Isomerase; Membrane; Repeat.
 FT REPEAT 69 110 PEST 1.
 FT REPEAT 418 459 PEST 2.
 FT REPEAT 486 526 PEST 3.
 FT REPEAT 534 584 PEST 4.
 SQ SEQUENCE 658 AA; 74068 MW; 26B6AC3C1BD625C CRC64;

alignment_scores: Quality: 94.50 Length: 206
 Ratio: 1.050 Gaps: 9
 Percent Similarity: 43.689 Percent Identity: 22.330

alignment_block:
 US-09-303-518D-569 x SOHC_ZYMMO ..

Align seg 1/1 to: SOHC_ZYMMO from: 1 to: 658

181 ATGCTCTGCGCCGACGACGTTC...CCCAAAAGTGCAAGCCGT 227
 ::::: ||| ::::: |||
 106 LeuLysMetIleGlyAspSerProAspAlaProHisMetLeuArgAlaAr 122
 228 TTTTCGGGAAAGCGCAAAAGCGGTTTGAACTGCCCGCGGTTTTTCA 277
 |||||
 122 gAsnGluLeuAlaArgGlyAlaMetArgAlaAsnValPheThrA 139
 278 GAAACCGGAAACATAGAAACATGTTCAAAAGCGGTACAGCGGTGGAA 327
 |||
 139 rg.....IleGlnLeuAlaLeuPheGlyAlaMetSerTrpGlu 151
 328 CATGTGACAGCGCTTTGGACAAACACGAGGCGTGTATTTCATCAGCC 377
 |||||
 152 HisValProGlnMetProValGlu.....LeuMetLeuMetPr 164
 378 G.....CAGATCGGCGCTACGATTGGCGGAGCGCTACA 412
 |
 164 oGluTrpPheProValHisIleAsnLysMet.....AlaTrpT 177
 413 TCAGCCAGCAGCTCCGTCCTCCGCTGACCGCCCATGTACAAACGCCGAA 462
 ::::: |||||
 177 rPalaArgThrValLeuValProLeuLeuVal..... 187
 463 ATCAAGCGATGACAAATCATGACGGCGGCGGTTCGGGCAAGG 512
 187 187
 513 AAAACCGCGCTACACGATCAAGGGGTCAACAAATCATCAAGGCC 562
 ::::: |||||
 188LeuGlnAlaLeuLysProValAlaArgAsnA 198
 563 TGGCTTCGGGCGAAGCA...ACCATGCTCTGCCCGACACGTCCTCC 609
 |||||
 198 rgArGlyIleLeuValAlaPGLuLeuPheValProAspValLeuProThr 214
 610 CTCGCAAGAGCGGGGAGCGGTATGGTGGATTCTTGGGC..... 651
 |||||
 215 LeuGlnLysSerLysProIleTrpArgPhePheSerAlaLeuAs 231
 652AACTGCTATACCATGACGCTGGCGGCA 682
 231 pLysValLeuHisLysValGluProLysTrpProLysAsnMetArgAla 248
 683 AATTGGCACACGTCAAAAGCGGTGAACCCGTTTCTGCTCGCAAGCC 732
 |||
 248 ySaLalIleHis.....SerCysValHisPheValThrGluArg 260
 733 CTGCTGCGCGGACAGGT 750
 ||| |||
 261 LeuAsnGlyGluAspGly 266

seq_name: SwissProt_40:YT44_STRFR

seq_documentation_block: PRT; 395 AA.
 ID YT44_STRFR STANDARD;
 AC P20188;
 DT 01-FEB-1991 (rel. 17, Created)
 DT 01-FEB-1991 (rel. 17, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Hypothetical 44.4 kDa protein in transposon TN4556.
 DE Streptomyces fradiae.
 OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1906;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TRANPOSON-Tn4556;
 RX MEDLINE=90185236; PubMed=2155856;
 RA Siemieniak D.R., Slightom J.L., Chung S.T.;
 RT "Nucleotide sequence of Streptomyces fradiae transposable element
 Tn4556: a class-II transposon related to Tn3.";
 RL Gene 86:1-9(1990).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M29297; AAA8562.1; -
 DR PIR; J0430; J0430.
 KW Hypothetical protein; Transposable element.
 SQ SEQUENCE 395 AA: 44379 MW: C50B556F0E3B2838 CRC64;

alignment_scores:
 Quality: 94.00 Length: 294
 Ratio: 0.783 Gaps: 13
 Percent Similarity: 40.816 Percent Identity: 22.109

alignment_block:

US-09-303-518d-569 x YT44_STRFR

Align seg 1/1 to: YT44_STRFR from: 1 to: 395

```

51 CATCTGTTGACGGCCCTGCTCAAAATGCTCTCCCTGCGCCGCTTTCCT 100
   ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 HISHIALASPGLNGLNLEUGLNARGALASPGLYALGLNPROVA 67
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 GTCTGCACAGCTGGGAAACCGGCTGACATCTGCTGTTTACCTTTA 150
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
67 IASPLNHSGLY.....ProProCysLeuValGlyIle..... 78
151 MAGGAAGACCGCGCGGCATCTGCGCAATATGCTCAGGACGATGAA 200
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 .....ArgValGlnArgArgGlyGlnGlnGlnGlnArgGlyAsp 91
201 TCCGACCCCAAAACGGTCAAAACGGTTCGGAACGCAAGCAAGCGG 250
   ||| ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
92 ARGARGGLYGLNLEUVALPROGLYHSHISLEUASPGLYARG..... 106
251 GTTGGAACTGCCCCCGCTTTTCAGAAAACCGAAGACATAGAACAA 300
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
107 .....ProHisProPheGlnAlaLeuHisPheArgArgArg 119
301 ATGTTCAAAAGCGGTACAGCGCTGGGAACATGTGCAGCAGGCTTT..... 344
119 SPLeu...ProGlyValGlyValAspLeuValAspGlyValGlnHis 134
345 .....GACAAACAGCAAGGGGTGCTATTCATCAGCCCGC 379
135 ASPLeuGlnArgGlyArgGlnLeuArgArgProLeuHisAlaArgG1 151
380 ACATCGGACGCTAGCATTTGGCGGACGCTACATCAGCAGCAGCTTCG 429
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
151 nArgLeuGlnLeuLeuArgAlaLeuAlaLeuArgProGlyGlnGln 168
430 TTTCCGCGTACCGCATGTACAAACCGCGCAA.....AATCAAGC 470
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
168 eArgArgArgPGLNHSLeuGlnGlnValGlnArgValAlaHisArgPro 184
471 GATACAAATCATGACAGCGGCGGCTTCGCGCAAGAAAGAAAACCG 520
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 ArgArgGlnValHisGlyGlyGlnGlnArgArgGln..... 197
521 CGCTACCAAGATACAGGGGTCAAAACAAATCATCAAGCCCTGCTTCG 570

```

```

198 .....A 198
571 GCGGAGCAACATCTGCTGCCGACATGCTCCCTC..... 608
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
198 lAlaSerValGlyValProAlaGlnGlnArgArgLeuAlaHisPro 214
609 .....CCCTCAGACGCGGGAAGCGGTATGCG 637
215 ValProGlyGlnLeuArgGlnProAlaArgArgTyrProArgHisLea 231
638 TGGATTCTTCGGAACCTGCTTACCATGATGCGCGGCAAAATG 687
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
231 nGly.....AlaHisSerGlnArgGlyAspLeuVal 242
688 GCACAGCTCAAGCGGTCAAAACCTGTTTCTGCTGCAACGCTGCC 737
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242 rProLeuGlnArg.....ProLeuHisLeu..... 250
738 TGGCGGACAGGTTTGCATTCACATCGCCGCCCTCAGAGGATTTGA 787
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
251 .....ArgThrArgHisProHisArgProAlaProGlnProValG1 264
788 ACGGCGCAAAACCGCATGATGCGCGCTGTTCAACCG..... 824
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 nLeuArg...AspProLeuProGlyArgHisArgProAlaGlyGln 280
825 .....CAATGCCGAATATGATACGCGCG 848
280 eThrAlaArgCysProValLeuSerIlePro 290

```

seq_name: SwissProt_40:GLT3_WHEAT

seq_documentation_block:

```

ID GLT3_WHEAT STANDARD: PRT: 660 AA.
AC P08488;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutenin, high molecular weight subunit 12 precursor.
OS Triticum aestivum (Wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CHINESE SPRING;
RX MEDLINE=86041862; PubMed=3840588;
RA Thompson R.D., Bartels D., Harberd N.P.;
RT "Nucleotide sequence of a gene from chromosome 1D of wheat encoding a
RL HMW-glutenin subunit.";
CC Nucleic Acids Res. 13:6833-6846(1985).
CC -!- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE
CC VISO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC -!- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
CC -!- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
CC GROUP 1 CHROMOSOMES OF WHEAT.
CC -!- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE
CC NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQDPGQ AND
CC GQDPGQGGQGYTPS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X03041; CA26847.1; -
CC DR PIR; A24266; A24266.
CC DR HSP; P01088; IBFA.

```

DR InterPro: IPR001768; Cereals tryp. amy1_inh.
 DR InterPro: IPR001419; Glutinin.
 DR Pfam: PF00234; tryp_alpha_amy1. 1.
 DR PRINTS: PR00210; GLUTENIN.
 KW Seed storage protein: Repeat; Multigene family; signal.
 FT SIGNAL 1 21
 FT CHAIN 22 660
 FT DOMAIN 125 615
 FT REPEATS 12.
 SEQUENCE 660 AA; 70868 MW; 2BFD09D8C8FCCCF CRC64;

alignment_scores:
 Quality: 94.00 Length: 160
 Ratio: 1.068 Gaps: 5
 Percent Similarity: 55.000 Percent Identity: 24.375

alignment_block:
 US-09-303-518D-569 x GLUT3_WHEAT ..

Align seg 1/1 to: GLUT3_WHEAT from: 1 to: 660

```

150 AAGAGAGAGCCGCGCATGTCGCCAATATGCGTCAGGAGCATGA 199
    ::::: ||| ::::: ||| ::::: ||| ::::: |||
319 GlnGlnHsTYrPrOAlaSerGlnGlnInPrOgLnGlnGlnGln 335
    200 ATCCGAGCCCAAAAGCGTCAAGCCGTTTGGGGAACGGCAAAAGCC 249
        ||| ::::: ||| ::::: ||| ::::: |||
335 YHsTYrPrOAlaSerGlnGlnInPrOgLnGlnGlnGlnGln 349
    250 GGTGGTAAGTGGCCCGCGTTCAGAAACCGGAAGACATAGAAC 299
        ||| ||| ||| ||| ||| ||| ||| |||
350 .... GlnGlnInPrOAlaSerGlnGlnInPrOgLnGlnGln 364
    300 AATGTTCAAGCGGTACAGCGGTGGAACATGTGACAGCGCTTGACA 349
        ||| ||| ||| ||| ||| ||| ||| |||
365 ..... GlnHsTYrPrOAlaSerLeuGlnGln 373
    350 AACGAGAGGCGTGTATTCATCAGCGGCACATGCGGAGTACGATTG 399
        | ::::: ||| ::::: ||| ::::: |||
373 nPrOgLnGlnGlnGlnHsTYrPrOThSerLeuGlnGlnGlnGln 390
    400 GCGGAGCGTACATCAGCAGCGATTCGCTCCGCTGACCGCCATGTA 449
        ||::: ||| ||| ||| ||| ||| ||| |||
390 YGlnGlnInLeuGlnInPrOgLnGlnGlnGlnGlnGlnGln 406
    450 CAACCGCCGGAATTCAGGATGACAAATCATGCAAGCGGCGGAGG 499
        ||||| ||| ||| ||| ||| ||| ||| |||
407 GlnThGlnGlnGlnGlnInPrOgLnGlnGlnGlnGlnGlnGln 423
    500 T.....TCGCGGCAAGGAAAAACCGCGCTACG 528
        | ||| ||| ||| ||| ||| ||| |||
423 YGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 440
    529 AGCATACAGGGGTCA.....ACAATCATCAAGCCCT 563
        ||::: ||| ||| ||| ||| ||| ||| |||
440 InGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 456
    564 GCGTGGGCGAAGCAACATGCTGCTGCC 593
        ::: ||| ||| ||| ||| ||| ||| |||
457 YGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 466
```

seq_name: SwissProt_40:FLA_MOUSE

seq_documentation_block:

ID FLA_MOUSE STANDARD: PRT; 336 AA.
 AC P11088-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Filaggrin (Fragment).
 GN FIG.
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88058903; PubMed=3680218;
 RA Rochnagel J.A., Mehrel T., Idler W.W., Roop D.R., Stehnel P.M.;
 RT "The gene for mouse epidermal filaggrin precursor, its partial
 RT characterization, expression, and sequence of a repeating filaggrin
 RT unit.";
 RL J. Biol. Chem. 262:15643-15648 (1987).
 RN [2]
 RP REVISIONS.
 RA Rochnagel J.A.;
 RL Submitted (SEP-1988) to the EMBL/GenBank/DBJ databases.

-1- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES
 CC DISULFIDE-BOND FORMATION AMONG THE INTERMEDIATE FILAMENTS DURING
 CC TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
 CC -1- PTM: FILAGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
 CC HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES
 CC OF 248 AA, WHICH ARE NOT SEPARATED BY "LARGE LINKER". THE
 CC PRECURSOR IS DEPOSITED AS KERATOHALIN GRANULES. DURING TERMINAL
 CC DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.
 CC

- This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC

DR EMBL: J03458; AAF5559.1; ..
 DR PIR: A28444; A28444.
 DR MGD: MGI:95553; F19.
 KW Phosphorylation; Developmental protein.
 FT NON_TER 1
 FT SEQUENCE 336 AA; 35678 MW; 259F124D3AC0DB2D CRC64;

alignment_scores:
 Quality: 93.50 Length: 176
 Ratio: 1.140 Gaps: 9
 Percent Similarity: 46.591 Percent Identity: 25.568

alignment_block:
 US-09-303-518D-569 x FLA_MOUSE ..

Align seg 1/1 to: FLA_MOUSE from: 1 to: 336

```

165 GCGCATGCTGCCAATATGGGTGAGGAGCATGATCCGACCCGAAA 214
    ||||| ||| ||| ||| ||| ||| ||| |||
128 AAlaHsArGlnSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 144
    215 CGTCAAGCGCGTTCGCGAAGCGC.....AAAAGCGGTTG 255
        ||||| ||| ||| ||| ||| ||| ||| |||
144 gGlnGlnHs.....GlySerGlnGlnGlnGlnGlnGlnGln 158
    256 GAACCTGGCCCCGCTTTTCAGAAACCGGAAGACATTAACAAATGTT 305
        ::: ||| ||| ||| ||| ||| ||| |||
158 ergLuglnGlnSerAspSerGlnHsGlnHsGlnHsGlnHsGln 174
    306 CAAGCGGTACAGCGGTGGGAACATGTGACAGAGCTTGGACAACACG 355
        ||| ||| ||| ||| ||| ||| ||| |||
175 GlnArGln..... 177
    356 AAGGCTGTCTATTCATCAGCGCGCATCGGACATTCGATTGGCGGGA 405
        ||::: ||| ||| ||| ||| ||| ||| |||
178 .....HsGlnHsGlnHsGln..... 183
    406 CGCTACATCAGCAGCAGCTTCGCTCCGCTGACCGGCATGTGACAAAC 455
        ||||| ||| ||| ||| ||| ||| ||| |||
184 .....HsGlnHsGlnHsGlnGlnGlnGlnGlnGlnGlnGlnGln 197
```

```

456 GCCGAATCAAGCATAGACAAATCATGACGGCGGCGTTGCG 505
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
198 .....GlnGlnSerSerGlyArgGlyHisGlnGlyAlaHisGlnGlu 212
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
506 GCAAGG.....AAAAACCGCGCTACCACGAA 534
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
212 nGlyArgAspSerAlaArgProArgGlySerAsnGlnGlyHisSerS 229
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
535 CAAGGGGTCAACCAATCAT..... 554
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
229 eArgHisGlnAlaAspSerProArgValSerAlaArgSerGlySerGly 245
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
555 .....CAAGCCCT...GGTTGCGGCGAGCAACCATGCTCTGC 592
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
246 GlyArgGlyGlnSerProAspAlaSerGlyArgSerSerAlaArgAs 262
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
593 CGGACGACGT...CCGCTCCCTCAGAA 617
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
262 pArgProArgGlnProSerProSerGln 271
      ::::::::::: ::::::::::: ::::::::::: :::::::::::

```

seq_name: SwissProt_40:YHL1_EBV

seq_documentation_block:

```

ID YHL1_EBV STANDARD: PRT; 660 AA.
AC P03181.
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical BHLP1 protein.
OS Epstein-Barr virus (strain B95-8) (human herpesvirus 4).
OC Viruses; dsDNA viruses; no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OK NCBI_taxonomy:10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tufnell P.S., Barrett B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V01555; ; NOT_ANNOTATED_CDS.
DR PIR: A03742; Q0BE3.
KW Hypothetical protein; Early protein; Repeat.
FT DOMAIN 149 648 4 x 125 AA RANDOM REPEATS.
FT REPEAT 149 273 1.
FT REPEAT 274 398 2.
FT REPEAT 399 523 3.
FT REPEAT 524 648 4.
SO SQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

```

alignment_scores:

```

Quality: 93.00 Length: 282
Ratio: 0.823 Gaps: 11
Percent Similarity: 40.071 Percent Identity: 25.177

```

alignment_block:

US-09-303-518D-569 x YHL1_EBV

Align seg 1/1 to: YHL1_EBV from: 1 to: 660

27 TCCCCCTTTGGAAACCGCATGACATCTGTTGACCGCGCTGCT..... 71

```

211 ThrProHisProGluArgGlySerGlyProAlaAspProProAlaAla 227
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
72 .....CAATGCGCTCTCCCTGTCGCG 93
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
227 aArgLeuProProGluArgGlnGluProArgLeuProGlnsPleuAla 244
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
94 CTTCCTGCTGTCGACACGCTGGGAAACCGGCTCGACATCTGGCGTTTA 143
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
244 IaAlaGlnArgCysProAlaGlyProProProThrArgSerGlyAla... 259
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
144 CTTTAAAGAGAGACCGCGCGCATGCTGCGCAATTTGGCTCAGGAG 193
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
260 .....AlaAlaGlnArgThrHisArgArgProProGlyCysProAr 273
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
194 GCATGAATCCGACCCCAAAACGGTCAAGCCGTTTGGCGAAAGGCA 243
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
273 gSerAlaArgAsnPro..... 278
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
244 AAAGCGGTTTGGAACTTGCCCGCGCTTTTTCAGAAAACCGAGACAT 293
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
279 .....GlyCysProArg.....ThrTrpArgArg 286
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
294 AGAAACATGTTCAA..... 308
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
287 ArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGlyAlaArgPr 303
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
309 .AGCGGTACAGCGGTGGGACATGTGCAGCAGGCTTTGGACAAACAGAA 357
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
303 oSerGlyProThrGlyArgProAlaAlaProGlyAlaProGlyThr 320
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
358 GGGGTGCATTCATCAGCGCGACATGCGACATGTCATTTGGCGGAGC 407
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
320 roAlaAlaProGlyProGlyGlyAlaAlaAlaProSerGlyAlaThr 336
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
408 CTACATCAGCCAGCAGCTCCGTCCTCCGCTGA .CCGCGATGTCAAACG 456
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
337 ProHisProGluArgGlySerGlyProAlaAspProProAlaAlaAla 353
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
457 CCGAAATCAAGAGATAGACAAATCATGACGCGGCGAGGTTCCGCG 506
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
353 gLeuProProGluArg..... 358
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
507 CAAGGAAACCGCGCTCTACGACATACAAAGGGTCAACCAATCATCA 556
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
359 .....GlnGluProArgLeuPro..... 364
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
557 AAGCCTGCGTGGCGGAGCAACATCGTCTGCGCCGACAGTCCCG 606
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
365 GlnAspLeuAlaAlaGlnArg.....CysProAlaGlyProPr 378
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
607 TCCCTCAAGAGCGGCGAGCGCTAT.....G 635
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
378 oProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProArg 395
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
636 GGTGCAATTTCTGCGCAACCTGCT...ATACATGACGCTGGCGGAA 682
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
395 LysProArgSerAlaArgAsnProGlyCysProArgThrTrpArgArg 411
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
683 AATTGCGACACGTCAAGGCGGTAAAAACCTGTTTCTGTCGCAACGC 732
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
412 Arg.SerGlyAlaGlnArgGlyHisProProProProGlyAlaGlyAlaArgP 428
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
733 CTGCTGGCGGACAGTTTGATTTGCGACATCCGCGCGCTGCA 776
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
428 roSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaPro 442
      ::::::::::: ::::::::::: ::::::::::: :::::::::::

```

seq_name: SwissProt_40:POLN_RRV1

seq_documentation_block:

```

ID POLN_RRV1 STANDARD: PRT; 1148 AA.
AC P13888;
DT 01-JAN-1990 (Rel. 13, Created)

```


alignment_scores: Quality: 92.50 Length: 281
 Ratio: 0.690 Gaps: 13
 Percent Similarity: 47.687 Percent Identity: 20.641

alignment_block:
 US-09-303-518D-569 x VP10_RDVF ..

Align seg 1/1 to: VP10_RDVF from: 1 to: 353

```

211 AAACGGTCAA...GCCGTTTGGGAAACGCAAAAGCGGTTTGA 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
32 LysLysValLysLeuGlyThrLeuAlaAsnGlnSerGlyAlaAsnAs 48
258 ACTTGCCCGCGGTTTTCAGAAACCGCAACACATAGAAACAATGTA 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
48 nlethrgluAlaPhePheAspLysLeuArgAsnGlnArgLysSerG 65
308 AAGCGGTACAGCGGTGGGAACATGTGACACAGGCTTG...GACAAA 351
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 luAlaTyrLeuAlaSerAspLeuAlaGluArgGluLeuThrArgSp 81
352 CACGAAGGCTGCTATCATCATCGCCGACATCGGACGCTACGATTGG 401
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
82 HisLysAlaIleValPheValThrLys.....SerValLeuLeuG 95
402 CGAGCGCTACATCAGCAGCAGCAGCTTCGCTCCGCTGACCGCC... 444
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 yglLysSerLeuLysAspLeuLeuProTyrGlyValIleValGlyAla 112
445 ..ATGTACAAACCGCGCAAAATCAAGCAGATAGCAAAATCATGAGC 492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
112 heilePheIleProGluThrAlaSerValLeuAspAsnVal..... 125
493 GCGAGGTTCCGCGCAAGAAACCGCGCTACACAGATACAGGGGT 542
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
126 ..ArgValMetIleGlyAsnGlnLysArgProLeuThrValAlaLeu 141
543 CAACAAATCATCAAAAGCCCTGCGTTCG.....GlyAspSerTyrA 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
141 elystrMetAlaLysSerLeuAsnCysAspLeuValGlyAspSerTyrA 158
571 .....GGGAAAGCAACCATC... 585
158 spThPheTyrTyrCysAsnSerSerAlaTyrGlyLysAsnLeuLeu 174
586 GTCTGCGCCGACCAAGCTCCCTCCCTCAGAAAGCGGGAAGCGGTG 635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
175 ValSerGluAsnAspPheSerAsnProGlnArgAlaLeuLeuSerVal 190
636 GGTGATTTCTTCGGCAAACTGCTATACATGACGCTGGCGCAAAAT 685
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
191 .....GlyAspLeuCysTyrGln.....AlaAlaArgS 200
686 TGGCACAGCTCAAAAGCGTGAACCCGTTTTCGTGGGAAGCGCTG 735
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
200 erlleHisValAlaAlaAlaAsnTyrIleArgGlie...PheAspArgMet 215
736 CCGTGGGACAGAGGTTTGCATTCGACATCCGCCCGCTCAAGGGGAAT 785
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
216 ProGlyPheGlnProSerLysHisLeuPheArgIleIleGlyValLe 232
786 GAAC..... 789
232 uasPheIleuThrLeuLysThrMetValThrSerAsnIleAlaArgLup 249
790 .....GGCGACAAGGCCCATGATGCGCGC 813
249 roglYmetPheSerHisAspAsnValLysAspValLeuHisArgThrGly 265
814 GTGTTCACCGCAATGCCGAATAT.....TGGATACCGCG 848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

266 ValPheSerProAsnHisHisPheSerAlaValIleLeuTpr...ArgGI 281
 849 TTTTCCGACGCGATCTGTTATGTACACCGCTACAAATG 891
 281 yTrpAlaSerThrTyrAlaTyrMetPheAsnGlnGluGlnLeu 295

seq_name: SwissProt_40:SPR6_HUMAN

seq_documentation_block:
 ID SPR6_HUMAN STANDARD; PRT; 344 AA.
 AC Q13247; Q13244; Q13245; Q9UJB8;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Splicing factor, arginine/serine-rich 6 (Pre-mRNA splicing factor
 DE SPR55).
 GN SPR6 OR SPR55.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE=Colon;
 RX MEDLINE=96016206; PubMed=7556075;
 RA Sreelton G.R., Caceres J.F., Mayeda A., Bell M.V., Plebanski M.,
 RA Jackson D.G., Bell J.I., Krainer A.R.,
 RT "Identification and characterization of three members of the human SR
 RT family of pre-mRNA splicing factors".
 RL EMBO J. 14:4336-4349(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stuurges G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.D.,
 RA Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grahnam D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.W., King A., Knights A., Laird G.K., Lawlor S.,
 RA Levaslainho M.H., Leverhwa M., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McComachie L.J., McElay K., McMurtry A.A.,
 RA Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prichaltingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilting L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RN [3]
 RP SEQUENCE OF 21-27 AND 47-55.
 RX MEDLINE=92249775; PubMed=1577277;
 RA Zahler A.M., Lane W.S., Stolk J.A., Roth M.B.;
 RT "SR proteins: a conserved family of pre-mRNA splicing factors.";
 RL Genes Dev. 6:837-847(1992).
 CC -1- FUNCTION: PLAYS A ROLE IN CONSTITUTIVE SPLICING AND CAN MODULATE
 CC THE SELECTION OF ALTERNATIVE SPLICING SITES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -2- ALTERNATIVE PRODUCTS: 3 ISOFORMS; SPR55-1 (SHOWN HERE), SPR55-
 CC 2 AND SPR55-3; ARE PRODUCED BY ALTERNATIVE SPLICING AND DIFFER IN
 CC THEIR C-TERMINAL DOMAINS.
 CC -1- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS
 CC DOMAIN (BY SIMILARITY).

1

Percent Similarity: 43.723 Percent Identity: 23.377

alignment_block:
US-09-303-518D-569 x VE2_HPV37 ..

Align seg 1/1 to: VE2_HPV37 from: 1 to: 454

```
273 TTTTCAGAAACCGGAGACATAGAAACATGTTCAAGCGGTACA.... 317
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
179 PheSerLysThrGlyArgTrpGluValHisValAsnLysAspThrIlePh 195
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
318 .....CGGCTGGGAACATGTGCAGACAGCTTTGGACAAACAGCAA 357
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
195 eAlaProValThrSerSerSerProAlaIleGluThrAspGlyA 212
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
358 GGGCTGCTATTATCATCCAGCCGACATCGGCAGCTA..... 392
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
212 IaIaSerValHisThrValSerGlySerProLeuAlaArgLysPheSer 228
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
393 .....CGATTGGCG 403
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
229 ThrThrSerValSerThrArgLysArgThrProArgArgTyrArgAr 245
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
404 GAGCCATACATCACCACAGCTCCGCTCCGCTGACCGCATGTACAAA 453
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
245 GlyAlaSerSerProThrThr.....AlaAlaArgGlnLysArgG 260
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
454 CCGCCGAATAATCAAGCATAGACAAATCATGCA..... 488
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
260 InglyGluAspThrAlaThrArgSerArgSerThrSerArgLys 276
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
489 .....GGCGGGCAGGGTCCGCGCAAGGAAACCGCGC 523
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
277 GlnAlaThrSerArgGlyGlyAspArgArgArgArgArgGlnArgSe 293
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
524 CTACCAG.....CATA 534
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
293 rTyrSerArgAspThrSerSerSerProAspArgLysArgGlyArgS 310
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
535 CAAGGGGT.....CAACAAATCATCAAGCCCTGGGTCCGGCGAGC 578
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
310 eArgGlyGlyProGluThrArgSerGlnSerArgSerLeuSerArgSer 326
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
579 AACCATGCTCCTGCCGACACGATCCCTCCCTCAAGAAAGCGGGGAAG 628
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
327 ArgSerArg.....SerArgSerArgGlySe 335
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
629 GCGTATGGGTGATTTCTTGGCAAACTGCTTACCATGACGCTGGCG 678
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
335 rSerSerArgGlyValAlaPro.....AspAlaValG 347
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
679 GCAAA.....ATTGCG..... 689
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
347 lYlYSerValArgThrValGlyArgAspHisSerGlyArgLeuLysArg 363
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
690 ...ACACGTCAAAAGCGTGAACCCCTTTTCTGCTGCGAAGCCCTGC 736
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
364 LeuLeuAspGluAlaArgAspProValIleValLeuArgGlyAspAl 380
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
737 C.....TGCGGACAAAGSTTTCGATTTCACAT 764
||||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
380 aAsnLysLeuLysCysTyrArgTyrArgAlaLysLysLysHis 394
```